CUK-H000663-14-294092

The British Journal for the Philosophy of Science

294092

14

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

Volume 23

CAMBRIDGE UNIVERSITY PRESS 1972

CONTENTS

ARTICLES, DISCUSSIONS, MISCELLANEOUS

ARTICLES

page

1 TURING MACHINES AND THE MIND–BODY PROBLEM
J. J. Clarke

13 QUANTUM-THEORETICAL REALISM: POPPER AND EINSTEIN V. KOCHEN AND SPECKER
Michael R. Gardner

25 EMPIRICAL SIMPLICITY AS TESTABILITY
Kenneth S. Friedman

35 THE PROPENSITY THEORY OF PROBABILITY
Alan R. White

89 TWO DEVIANT LOGICS FOR QUANTUM THEORY: BOHR AND REICHENBACH
Michael R. Gardner

111 THE INCOMPATIBILITY OF MACH'S PRINCIPLE AND THE PRINCIPLE OF EQUIVALENCE IN CURRENT GRAVITATION THEORY
James F. Woodward *and* Wolfgang Yourgrau

191 NUMBERS AND NUMERALS
William C. Kneale

207 SITUATIONAL DETERMINISM IN ECONOMICS
Spiro J. Latsis

295 NEWCOMB'S PARADOX REVISITED
M. Bar-Hillel *and* A. Margalit

305 BIOLOGICAL CLASSIFICATION
V. J. F. Pratt

DISCUSSIONS

45 QUINE'S PHILOSOPHICAL NATURALISM
J. Giedymin

55 EPICYCLIC POPPERISM
E. E. Harris

117 ON THE MACH PRINCIPLE AND RELATIVE SPACE-TIME
M. Sachs

119 RELATIVITY AND THE REALITY OF PAST AND FUTURE EVENTS
R. Weingard

247 METAPHYSICAL COMMON SENSE
J. L. Mackie

page
252 ONTOLOGICAL VARIANCE AND SCIENTIFIC OBJECTIVITY
M. Martin

329 TWO POINTS IN THE THEORY OF STATISTICAL INFERENCE
G. Barnard

331 PROPENSITY THEORIES OF PROBABILITY UNSCATHED: A REPLY TO WHITE
T. Settle

335 A COMMENT ON GILLIES'S FALSIFYING RULE FOR PROBABILITY STATEMENTS
J. De Maré

335 REPLY TO DE MARÉ
D. A. Gillies

336 FREE WILL IN A MECHANISTIC UNIVERSE? AN EXTENSION
D. A. Evans *and* R. T. Landsberg

343 I'M FREE BECAUSE I KNOW THAT I DON'T YET KNOW WHAT I'M GOING TO DO?
J. McDermott

REVIEW ARTICLES

69 BACK TO ARISTOTLE? (review of R. Harré *The Principles of Scientific Thinking*)
D. Miller

123 THE LOGIC OF STATISTICAL INFERENCE (review of I. Hacking *The Logic of Statistical Inference*)
G. A. Barnard

132 LIKELIHOOD (review of A. F. Edwards *Likelihood. An Account of the Statistical Concept of Likelihood* and *its Application to Scientific Inference*)
I. Hacking

138 THE SUBJECTIVE THEORY OF PROBABILITY (review of de Finetti *Teoria delle Probabilità, Sintesi Introduttiva con Appendice Critica*)
D. A. Gillies

157 THE PLAIN MAN'S GUIDE TO PROBABILITY (review of J. R. Lucas *The Concept of Probability*)
C. Howson

170 THE SIGNIFICANCE TEST CONTROVERSY (review of D. E. Morrison and R. E. Henkel (*eds.*) *The Significance Test Controversy: A Reader: Methodological Perspectives*, E. F. Heerman and L. A. Braskamp (*eds.*) *Readings in Statistics for the Behavioural Sciences*, B. Lieberman (*ed.*) *Contemporary Problems in Statistics*, J. A. Sleger (*ed.*) *Readings in Statistics for the Behavioural Scientist* and E. R. Tufte (*ed.*) *The Quantitative Analysis of Social Problems*)
R. N. Giere

181 BAYESIANISM AND THE RATIONALITY OF SCIENTIFIC INFERENCE (review of R. G. Colodny (*ed.*) *Essays in Contemporary Science and Philosophy:* volume 3, *Mind and Cosmos;* volume 4, *The Nature and Function of Scientific Theories*)
J. Dorling

page

257 HEGEL'S PHILOSOPHY OF NATURE (review of G. W. F. Hegel *Philosophy of Nature*)
G. Buchdahl

266 SCIENCE, HISTORY AND METHODOLOGY (review of I. Lakatos and A. Musgrave (*eds.*) *Criticism and the Growth of Knowledge*)
J. J. C. Smart

274 FOR AND AGAINST METHOD (review of M. Radner and S. Winokur (*eds.*) *Analyses and Methods of Physics and Psychology*)
N. Koertge

347 LOUIS PASTEUR—A CASE AGAINST REDUCTIONIST HISTORIOGRAPHY (review of R. Dubos *Louis Pasteur, Free Lance of Science* and F. Dagognet *Méthodes et Doctrines dans l'Oeuvre de Pasteur*)
N. Roll-Hansen

REVIEWS

BUCHDAHL, G. *Metaphysics and the Philosophy of Science*, review by D. J. O'Connor, 79

BUTTS, R. E. and DAVIS, J. W. (*eds.*): *The Methodological Heritage of Newton*, review by A. R. Hall, 80

ERWIN, E. *The Concept of Meaninglessness*, review by J. W. N. Watkins, 83

GUNDERSON, K. *Mentality and Machines*, review by J. L. Cohen, 292

KÖRNER, S. *What is Philosophy?*, review by H. Skolimowski, 366

SPENCER BROWN, G. *Laws of Form*, review by L. L. Whyte, 291

STEEN, S. W. P. *Mathematical Logic: with Special Reference to the Natural Numbers*, review by J. L. Bell, 363

MISCELLANEOUS

85–88, 370–373 RECENT PUBLICATIONS

BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE
374 Programme of Meetings 1972–73

CENTER FOR THE PHILOSOPHY OF SCIENCE, UNIVERSITY OF PITTSBURGH
374 Thirteenth Lecture Series, 1972–73

BOSTON COLLOQUIUM FOR THE PHILOSOPHY OF SCIENCE
375 Thirteenth Annual Programme

INDEX OF AUTHORS

ARTICLES

Bar-Hillel, M. (M. Bar-Hillel and A. Margalit) Newcomb's Paradox Revisited, 295

Clarke, J. J. Turing Machines and the Mind–Body Problem, 1

Friedman, K. S. Empirical Simplicity as Testability, 25

Gardner, M. R. Quantum-Theoretical Realism: Popper and Einstein v. Kochen and Specker, 13

Gardner, M. R. Two Deviant Logics for Quantum Theory: Bohr and Reichenbach, 89

Kneale, W. C. Numbers and Numerals, 191

Latsis, S. J. Situational Determinism in Economics, 207

Margalit, A. (M. Bar-Hillel and A. Margalit) Newcomb's Paradox Revisited, 295

Pratt, V. J. F. Biological Classification, 305

White, A. R. The Propensity Theory of Probability, 35

Woodward, J. F. (J. F. Woodward and W. Yourgrau) The Incompatibility of Mach's Principle and the Principle of Equivalence in Current Gravitation Theory, 111

Yourgrau, W. (J. F. Woodward and W. Yourgrau) The Incompatibility of Mach's Principle and the Principle of Equivalence in Current Gravitation Theory 111

DISCUSSIONS

Barnard, G. Two Points in the Theory of Statistical Inference, 329

De Maré, J. A Comment on Gillies's Falsifying Rule for Probability Statements, 335

Evans, D. A. (D. A. Evans and R. T. Landsberg) Free Will in a Mechanistic Universe? An Extension, 336

Giedymin, J. Quine's Philosophical Naturalism, 45

Gillies, D. A. Reply to De Maré, 335

Harris, E. E. Epicyclic Popperism, 55

Landsberg, R. T. (D. A. Evans and R. T. Landsberg) Free Will in a Mechanistic Universe? An Extension, 336

Mackie, J. L. Metaphysical Common Sense, 247

Martin, M. Ontological Variance and Scientific Objectivity, 252

McDermott, J. I'm Free Because I Know that I Don't Yet Know What I'm Going to Do? 343

Sachs, M. On the Mach Principle and Relative Space-Time, 117

Settle, T. Propensity Theories of Probability Unscathed: A Reply to White, 331

REVIEW ARTICLES

Barnard, G. A. The Logic of Statistical Inference (review of R. Harré *The Principles of Scientific Thinking*), 123

Buchdahl, G. Hegel's Philosophy of Nature (review of G. W. F. Hegel *Philosophy of Nature*), 257

Dorling, J. Bayesianism and the Rationality of Scientific Inference (review of R. G. Colodny (*ed.*) *Essays in Contemporary Science and Philosophy:* volume 3, *Mind and Cosmos;* volume 4, *The Nature and Function of Scientific Theories*), 181

Giere, R. N. The Significance Test Controversy (review of D. E. Morrison and R. E. Henkel (*eds.*) *The Significance Test Controversy: A Reader: Methodological Perspectives*, E. F. Heerman and L. A. Braskamp (*eds.*) *Readings in Statistics for the Behavioural Sciences*, B. Lieberman (*ed.*) *Contemporary Problems in Statistics*, J. A. Sleger (*ed.*) *Readings in Statistics for the Behavioural Scientist* and E. R. Tufte (*ed.*) *The Quantitative Analysis of Social Problems*), 170

Gillies, D. A. The Subjective Theory of Probability (review of de Finetti *Teoria delle Probabilita, Sintesi Introduttiva con Appendice Critica*), 138

Hacking, I. Likelihood (review of A. F. Edwards *Likelihood. An Account of the Statistical Concept of Likelihood and its Application to Scientific Inference*), 132

Howson, C. The Plain Man's Guide to Probability (review of J. R. Lucas *The Concept of Probability*), 157

Koertge, N. For and Against Method (review of M. Radner and S. Winokur (*eds.*) *Analyses and Methods of Physics and Psychology*), 274

Miller, D. Back to Aristotle? (review of R. Harré *The Principles of Scientific Thinking*), 69

Roll-Hansen, N. Louis Pasteur—A Case Against Reductionist Historiography (review of R. Dubos *Louis Pasteur, Free Lance of Science* and F. Dagognet *Méthodes et Doctrines dans l'Oeuvre de Pasteur*), 347

REVIEWS

Bell, J. L.	S. W. P. Steen: *Mathematical Logic: with Special Reference to the Natural Numbers*, 363
Cohen, L. J.	K. Gunderson: *Mentality and Machines*, 292
Hall, A. R.	R. E. Butts and J. W. Davis (*eds.*): *The Methodological Heritage of Newton*, 80
O'Connor, D. J.	G. Buchdahl: *Metaphysics and the Philosophy of Science*, 79
Skolimowski, H.	S. Körner: *What is Philosophy?*, 366
Watkins, J. W. N.	E. Erwin: *The Concept of Meaninglessness*, 83
Whyte, L. L.	G. Spencer Brown: *Laws of Form*, 291

Brit. J. Phil. Sci. 23 (1972), 1–12 *Printed in Great Britain*

501·05
B774
Vol. 23 Nos. 1–4
Feb – Nov. 1972

Turing Machines and the Mind-Body Problem

by J. J. CLARKE

1 In his [1961] Hilary Putnam has argued that 'all the issues and puzzles that make up the traditional mind-body problem . . . arise in connection with any computing system capable of answering questions about its own structure'. This shows, he believes, that such issues and puzzles are wholly linguistic or logical in character, and 'have thus nothing to do with the unique nature (if it *is* unique) of human subjective experience'.[1]

His argument makes use of the notion of a *Turing machine.* Such a machine consists of (*a*) a physical tape, scanner and printing device, along with a control and storage mechanism, and (*b*) a machine table which comprises a set of instructions for the machine to follow, and which directs the machine to print, erase or move the tape as required. There are thus two kinds of descriptions which can be given of a Turing Machine, the one called its *logical* description which concerns only the table of rules and instructions and does not presuppose any specific physical embodiment (it may be electronic relays, clerks sitting at desks, or whatever), and the other called its *structural* or *physical* description which applies to the specific system chosen to carry out the instructions of the machine table. These two types of descriptions, Putnam goes on to note, are closely analogous to the two types of descriptions typically applied to human beings, namely the *psychological* and the *physical.* Thus, the physical description of a human being corresponds to the physical (or structural) description of the machine, and the psychological description corresponds to the logical.

A few words of explanation are needed for the second half of this analogy. Putnam makes the following three points in this connection:

(1) The functional organisation (problem-solving, thinking) of the human being or Turing machine can be described in terms of the sequences of mental or logical states respectively (and the accompanying verbalisations), without reference to the "physical realisation" of these states.
(2) The states seem intimately connected with *verbalisation.*
(3) In the case of rational thought (or computing), the "programme" which

Received 19 *December* 1970

---

[1] For a further development of the argument, see also his [1967] and his [1964].

determines which state follows which, etc., is open to rational criticism (*op. cit.* p. 149).

Once it is observed that there is a close analogy between the two kinds of descriptions of a Turing machine on the one hand, and the two kinds of descriptions of human beings on the other, it becomes apparent that the whole question of the identity or non-identity of the mental and the physical is a bogus one. For who in his right mind would treat seriously the question of whether the physical and logical states of a machine are identical? Commitment to human souls, separate and distinct from their bodies, means commitment also to Turing machine souls, and if such questions are absurd in the case of machines, they must *mutatis mutandis* be absurd in the case of human beings to which they are closely analogous in the relevant respects. Thus, a machine may be in 'state A', which is defined in terms of the machine table, and, in the case of this particular machine, being in state A corresponds to the physical state where, say, flip-flop 36 is on. But there is no question here of 'state A' being or not being a different thing or process from being in the state where flip-flop 36 is on. To raise the question at all is to see its absurdity, to see that it is an empty verbal quibble, and in view of the analogy already drawn the absurdity accrues equally to the mind-body problem as well.

The primary purpose of this paper is to show that Putnam's *via analogica* does not offer any escape route from 'the issues and puzzles which make up the traditional mind-body problem'. The examination of the supposed analogy will also throw some interesting light on the nature of thinking by way of pointing a sharp contrast between the latter and the logical processes of Turing machines.

2 Putnam maintains that 'a "machine table" *describes* a machine if the machine has internal states corresponding to the columns of the table' (*ibid.* p. 141; his italics). This is certainly a requirement of his argument, for he develops this into the claim that there are two sorts of descriptions which can be given of a machine, the logical and the physical, following which the analogy is drawn with the mental and physical descriptions of human persons. What Putnam does not take account of is that *there is an indefinite number of possible self-consistent, but mutually exclusive, logical descriptions of a Turing machine which is in any given physical state or going through any given physical process.*

Let us for the sake of simplicity examine this claim with respect to a general purpose computer of the binary digital variety.[1] It will be remembered that the Turing machine as described in Turing's classic paper, "Computing Machinery and Intelligence" (*Mind*, 1950), is a "universal machine" in the sense that it can carry out in the simplest logical form any possible computation, and that therefore a Turing machine can in principle do anything that any computing machine can do. But in breaking down the process of computing into a series of atomic acts carried out on an infinite tape by means of a sparse unary notion, Turing has gained logical elegance at the expense of intuitive simplicity, and in presenting my objections to Putnam, therefore, I shall make use of the more easily manipulated idea of a general purpose computer.

---

[1] To make the analogy with human thinking processes as close as possible it would be necessary to postulate a mechanism with negative feedback.

For the moment the central processing and memory units of the computer can be considered as a Black Box; that is to say we do not need to know the manner in which it operates upon and stores its material.[1] In addition to this our computer has input and output units which, let us suppose, operate with Hollerith punch cards in which holes are punched in accordance with a code. The input cards, suitably perforated according to this code, carry instructions to the computer, such as to add or multiply two numbers, and the answer appears similarly encoded on the output card. In order to carry out more complex procedures than simple arithmetical computation, a set of systematically ordered instructions known as the "programme" is required, and these too are punched onto the cards by means of the code. Such programmes can consist either of complex numerical computations, such as calculating the mean of a set of numbers, or else by means of translations rules (i.e. another sort of code) of *non*-numerical procedures such as simulating a battle or translating from one language to another.

We can now apply Putnam's analogy to the general purpose computer. Its *logical* description will comprise its programme along with the coding rules for translating instructions into the "language" of punch cards. Its physical description will cover the Black Box and the perforated input/output cards. (We could of course give it "sense organs", but I will discuss this later.) Looking at the matter in this way, on what grounds can it be claimed that there is an indefinite number of possible true logical descriptions of a Turing machine which is passing through any given set of physical states?

To answer this we will first look at the matter from the general point of view of any code whatever. Given any coded message of finite length, and given a decoder of infinite powers who has no access to the rules of the code, the probability of any translation of the coded message approaches, but never equals, unity. This is intuitively obvious in the case of a relatively short message—say a dot-dash code with twenty "bits"—for here it is evident that the distribution of probabilities is insufficient to allow us to choose with any degree of assurance between a number of different possible self-consistent interpretations. As the length or number of the messages increases, however, the distribution pattern changes and becomes more stable and one interpretation emerges with a high degree of probability while the probability of its competitors diminishes. We normally decide that the translation with an overwhelmingly high probability is the "correct" translation, but this does not exclude in principle other possible self-consistent translations of the coded message, some of which may be entirely different semantically from the one we have chosen as "correct", or the possibility of several equiprobable translations.

---

[1] From the point of view of the human side of the analogy, the central nervous system is also treated as a Black Box.

We can apply this to the case of a computer. If its output consists of perforated punch cards, then such an output constitutes a code which requires translation rules to turn the message on the card into the language of arithmetic, or, by way of a more complex multi-stage de-coding procedure involved in the programme, into ordinary or technical language. Assuming for the time being that the programme is not known, then there is an indefinite number of possible self-consistent translations of the encoded output messages, however long or many such messages may be. Again, this is intuitively obvious in the case of a single punch card which, without much ingenuity, can with consistency be "made" to mean almost anything, but what appears as intuitive truth in the case of relatively short messages, applies equally to an encoded message of any length whatsoever.

The argument can be generalised to include the *complete* physical realisation of the machine table. Putnam has pointed out that a Turing machine can be embodied in a potentially infinite number of ways, and from any logical description thereof, no logical inference can be made as to its physico-chemical constitution (see his [1967], p. 186 ff.). Similarly, from the physico-chemical composition of our machine we cannot infer, as a matter of pure logic, its machine table. He believes that the latter can only be achieved if we are able to know 'how many significantly different states the machine is capable of and how these are causally related' ([1967], p. 187), but since we can never know for certain that we are in possession of a complete physical description of the machine, we can never know with certainty what such states might be. I on the contrary believe that even if *per impossibile* we were able to determine all future physical states of the machine, we would still not be in a position to make valid inferences concerning its logical states, for however large and complex such a set of physical states or processes may be there remains the possibility of alternative self-consistent interpretations. Again this is intuitively obvious in the case of a fairly simple set of processes, such as the activation of a single circuit which in itself could be made to mean anything at all. But it is equally true in the case of highly complex patterns of mechanical or electrical (or, for that matter, neuronal) activity, for, while logical consistency may require us to exclude a large number of possible interpretations, the number of such exclusions increasing proportionately to the complexity of the system, this number never narrows down to unity.

Furthermore, by means of a fairly simple procedure it would be possible to interpret any stack of punch cards as, for example, a musical score or a painting. I am not here referring to a computer which is actually *programmed* to write music or to paint pictures. The machine in question could be set to do certain computational tasks in arithmetic or programmed to navigate a space rocket, but nevertheless its output could be decoded as music or as painting. Let us imagine that a Martian lands on earth and in his tour of inspection comes across a functioning computer. If he is technically advanced enough to reach the earth it is highly likely that he will have made considerable advances in the development of computing machinery, and as a consequence it is also highly likely that he will recognise our computer for what it is. However, since he does not understand our language he will have no access to our codes or programmes except by means of an examination of the physical properties of the input and output cards or tape, but since he understands the general function of the apparatus and does not suppose that it is just a machine for producing fancy patterns,[1] one can imagine that he might send a stack of cards back to Mars for decoding. Now if the Martians are particularly interested in music and use their computers extensively for musical composition, they might, if they discerned an isomorphism between the patterns of perforated cards and their own musical grammar, come to the conclusion that what the earthman's computer is "really" doing is composing music.

Still, more fantastic, though still within the bounds of the possible, we can imagine that our Martians discover that the structure of the code is isomorphic with Martian *language* such that they can consistently transpose it into a piece of Martian prose. Of course the existence of such an isomorphism would be the rarest possible chance—somewhat of the same order as the production of the Bible by a million typing monkeys—but it is nevertheless conceivable. There would of course be no reason to suppose that the meanings of the individual words would be the same, for the computer has been set to perform tasks within an earthly, rather than a Martian, context. The Martian, then, could "read" the output of the computer quite consistently in terms of his own language, a reading which is opaque to its earthly operators.

---

[1] One can imagine, though, that a machine, which has been set merely to "doodle" in order, perhaps, for engineers to test its circuits, might produce "sense".

It might be objected that since the visiting Martians do not know the programme according to which the computer is operating, nor indeed the human purposes, plans and institutions which give our programmes the sense they have, then *his* isomorphic programme is merely parasitical. After all, he is in exactly the same position with respect to the physical states of our computers as he is with respect to the physical noises and marks produced by human language users. Now, as I have suggested, it is possible that our talk, simply as talk, has a physical structure which is isomorphic with Martian talk, but this would hardly be sufficient ground for saying that the *Martian*'s interpretation of what we mean when we talk is on all fours with what *we* claim to mean when we talk. And indeed if the Martians were to make a thorough examination of the "forms of life" in which our talk is embedded, they would eventually come to a correct interpretation of our language—namely the one which we ourselves would acknowledge.[1] So also with computers, the objection continues, for here too a sufficiently thorough understanding of the programme along with the "forms of life" from which it springs make it impossible to offer two or more equally possible interpretations of the computer's physical states.

This objection can be met in the following way. Under standard circumstances a computer will correctly be said to be performing one specific task, and one only, namely that which is specified by the programme currently being employed by the computer's operators. But it is perfectly possible in the case of isomorphic programmes that the computer may be performing two tasks simultaneously, not in the sense that the bits of each operation are occurring in different parts of the computer or in serial fashion, but rather in the sense that identically the same physical operations are interpretable as performing two quite distinct tasks. Let us imagine, for example, that a musician discovers that the programme for simulating an oil refinery can be translated with considerable effect into musical notation. He then proceeds to feed the instructions into the computer and to read off the output in terms of his own translation rules. At the same time the engineers from the oil refinery may be busy translating identically the same output cards in terms of production schedules. In such circumstances would it not be absurd to ask which of the two tasks the computer was "really" performing? Would it make sense to ask whether it is "really" running an oil refinery or composing music. It is surely quite arbitrary what sort of answer we give here.

---

[1] In principle, the Martian—assuming that he is a rational being—is in the same position as the anthropologist visiting a hitherto unknown tribe.

Thus, while it is true that in order to determine what a computer is doing (in the logical sense) we need to look beyond the purely physical structure of its output, to the whole human context in which it is operating, it is conceivable, nevertheless, that at one and the same time it may be straddling two or more quite different such contexts.

3 It is now time to see what are the implications of all this for Putnam's argument. To do this we must ask whether, just as there is an indefinite number of possible true descriptions of a Turing machine passing through a set of physical states, there is also an indefinite number of psychological descriptions of a person whose brain is passing through a set of physical states. The answer to this, I contend, is that there is not. Let us take the case of *thinking*, for it is here rather than in sensation or perception that Putnam's analogy appears most plausible. I shall take the term 'thinking' to refer, roughly, to "subjective" conscious episodes typically directed towards the solution of a problem or the making of a decision, or merely *about* something regardless of any purpose or outcome. We may now narrow down the problem and ask whether there is an indefinite number of possible true descriptions of what a person is *thinking* at any one time? There are two sorts of answer we can give to this, and we shall deal with each in turn.

In the first place it must be said that there *do* appear to be at any rate a *limited* range of possible descriptions of what I am thinking at any given time. Let us suppose that I am thinking about what to write next; I can describe this in many different ways, as thinking what the next sentence is to be, or how to follow on from the previous sentence, or how to continue the train of ideas, and so forth. In addition to this I could be more precise and describe the actual sentences and ideas with which I am toying. One's thoughts, in other words, can be described from various aspects and at various levels of generality in much the same way as physical occurrences, and in this sense it is evident that there is a range of possible descriptions of any given thought sequence.

However, at the same time it must be noted that there are significant limitations to the range of possible descriptions, limitations which do not apply to our descriptions of Turing machines. This limitation is imposed, roughly, by what might be called the 'subject matter' of the thought. In the set of thoughts which I have just been describing our range of possible descriptions is severely circumscribed by the rough area of discourse within which the thinking is taking place, and it is nonsense to suppose that the very same thoughts could be described or interpreted in terms of an entirely different area of discourse. In our previous discussion we saw, for example, that the output of a computer could consistently be interpreted or described in terms of arithmetic, ordinary language, music or painting, in other words in terms of four quite disparate areas of discourse. This feature is totally absent from our descriptions of thoughts, for to describe the above thoughts (the ones about what to write next) in terms of, say, arithmetic, is not to give an alternative description of that thought, but *to describe a totally different thought.*

We saw in the last section that arbitrariness may sometimes prevail when choosing how to describe a physical process. Thus, to use our previous example, it may in some circumstances be a matter of indifference whether we describe a computer as composing music or as working out production schedules. But no such arbitrariness or indifference prevails in the case of describing what a man is thinking. If I am giving a speech on how to run an oil refinery, a member of the audience, bored with the subject matter, though finding the *timbre* of my voice melodious, may endeavour to block out the meaning of my words and listen to my speech as pure music.[1] But how he or any other member of the audience cares to construe or interpret the sounds coming from my mouth has no bearing at all on the question of what I *think* I am saying or what I consciously intend to say. It should be clear from this that my argument does not rest on any epistemological assumptions about "direct access" or "incorrigibility" with respect to one's own thoughts, but rather on the fact that the kind of ambiguity or arbitrariness which may arise with physical processes does not arise with respect to persons and their thoughts.

But are not *overt actions* in the same category as thoughts in this respect? After all, for our visiting Martian a violin recital may appear as a complicated calisthenic exercise with the score viewed as a set of movement instructions, yet we do have criteria for rejecting this and any other deviant interpretation, criteria which have little to do with the thoughts going on in the mind of the performer. However it is not always the case that we can give an unequivocal answer to the question 'What is he (really) doing?' with respect to a given action, and sometimes we may have to allow several mutually exclusive though consistent interpretations. Thus, imagine that our Martian takes the violinist back with him to Mars to teach his fellow Martians a new form of calisthenics; what activity is the violinist *now* engaged in? It is clearly arbitrary what answer we give to this question. We may choose to say that he is *both* playing music *and* performing calisthenic exercises. Or we may choose to say that he is "really" playing music and that the Martians have simply misunderstood the whole thing. Or we may decide to say that, in view of the altered circumstances in which the "player" now finds himself, it is simply pedantic to insist that he is playing music at all. On the other hand we enjoy no such latitude with the question of what the violinist is *thinking* about when he is performing.

---

[1] In some Chinese dialects a single syllable may have as many as twelve "tones" which pick out totally different meanings. In such a case the resemblance to music is by no means absurd.

It may further be objected that my argument takes no account of the possibility of diverse descriptions of a particular thought process on the basis of some theory of *unconscious* thinking. It could be argued, for example, that just as dreams sometimes have a "latent" meaning which is different from their "manifest" meaning, so too with conscious thoughts, and if this is so then my attempt to draw a distinction between the mental and the physical fails. However, even if we grant that what we are "really" thinking sometimes differ from what we believe ourselves to be thinking, no threat is thereby posed to my argument, for according to such a theory it still remains true that at some level, whether conscious or unconscious, we are thinking what we are thinking regardless of any interpretations that may be adduced. Making the unconscious conscious may sometimes resemble a decoding procedure whereby the scraps and straws of conscious life are treated as clues to be pieced together into a consistent story or pattern, but there is no question here of arbitrary selection between equiprobable interpretations; only one interpretation can be accepted as the correct one. It has indeed been argued that the purpose of psychoanalysis is to place a consistent and therapeutically beneficial construction upon the relevant data rather than to uncover the "true" nature of the patient's motives, etc., but it is important to see that this is not, and cannot plausibly be, put forward as a theory of the *unconscious*, but rather as a theory about the nature of psychoanalysis.

We can now finally close in on our quarry. Why is it absurd to raise the question of whether, for example, a Turing machine's 'being in state A' is identical with the fact that 'flip-flop 36 is on'? The answer now seems obvious, and can be seen to be totally unrelated to questions about the identity or non-identity of the mental and physical processes of persons. Where logical descriptions of machines are concerned it is clearly nonsensical to wonder whether they describe the processes of souls or of mind-stuff distinct from their physical embodiment, for as we have seen an indefinite number of quite distinct logical descriptions can be given of one and the same set of physical processes. What the machine is "doing" in a logical sense is in certain given circumstances simply a matter of how we choose to interpret its activities. The absurdity lies in asking whether for each interpretation there is a separate soul or mental process which we elicit by means of our chosen interpretation. The human mind, on the other hand, is at any one time engaged in doing one and only one type of thing—working out a mathematical problem, composing a tune, or whatever—and it is thinking what it is thinking quite regardless of any interpretations we may want to place on its physical manifestations.[1]

It might be objected here that we have arbitrarily limited the notion of a Turing machine to the kind which is physically embodied in the form of a fairly simple machine such as is currently used for industrial and business purposes. But it may be assumed that the concept of a Turing machine covers not only these but also machines which are endowed with sensor and effector organs, and it may further be assumed that, at least as pencil and paper devices, such "machines" may come to resemble human beings very closely. And if all this is granted, then the analogy between human beings and Turing machines appears much less unlikely than I have suggested.

However it must be remembered that the analogy sought is between the logical *description* of a Turing machine and the mental *description* of a person, and not between machines and persons as such. If we permit ourselves to apply what Strawson has called P-predicates to an androidal Turing machine, then we may indeed have an analogy between Turing machines and human persons. And conversely, if we find it feasible to apply logical descriptions to human beings (for example in the event that the physical inputs and outputs of the central nervous system could be precisely computed), then once again we may rightly claim to be in possession of an analogy (though actually it is now beginning to look more like an identity than an analogy). However, neither of these analogies is of the kind which Putnam requires. For in effect Putnam is engaging in a sort of *reductionism.* He is trying to show that certain kinds of psychological descriptions operate in a closely analogous manner to the logical descriptions of Turing machines, as a consequence of which we can deal with the old mind-body problem without any reference to human beings at all. The hiatus that I have attempted to create, therefore, is not between Turing machines and human persons, but between on the one hand mental descriptions—whether applied to persons or Turing machines—and on the other hand logical descriptions—again whether applied to persons or to machines. Thus, our argument concerning the systematic ambiguity of a computer's physical processes could equally be applied to to the physical processes of a human person (say, to the EEG record), but we could hardly suppose that this is matched by any systematic ambiguity in what the person is actually thinking. Regardless, therefore, of whether other analogies may be drawn, this particular one cannot—not at any rate for the purpose for which Putnam requires it, namely as a quick way of retreat from a long embattled philosophical problem.

---

[1] Thinking two thoughts at the same time would be analogous to a computer carrying out two operations simultaneously in different parts of its machinery, and simultaneously issuing two sets of outputs.

4 We can now summarise the argument. Putnam wishes to postulate a close analogy between on the one hand the logical and physical states of Turing machines, and on the other the mental and physical states of human beings. By showing that with regard to Turing machines the mind-body problem is a will-o'-the-wisp, not worth a serious thought, he hopes to show that *mutatis mutandis* the traditional mind-body problem with regard to human beings is equally illusory. I have argued that on the contrary it is the analogy he seeks, rather than the problem he tries to dispel, which is illusory. Where fairly simple versions of Turing machines are concerned, e.g. ones with punch card input/outputs, the analogy with human thinking suffers from one crucial limitation, for whereas a person is thinking what he is thinking *simpliciter*, a Turing machine is "thinking" what it is "thinking" only by virtue of the manner in which we interpret its output or its internal states. Even the existence of a programme does not restore the missing piece of the analogy, for a computer may at any one time be operating in accordance with two distinct, though isomorphic, programmes. The addition of sense organs, limbs, etc., does indeed help to establish a close analogy—possibly even a (qualitative) identity—between machines and men, but this does not succeed in bridging the gap between mental and logical descriptions, whether they be of persons or of Turing machines.

The moral of this, I believe, is clear. There are limitations, crucial for the mind-body problem, on the extent to which our understanding of human persons is illuminated by the models and language of computer science. Unless we are prepared to embrace the view that each logical description of a Turing machine denotes a separate conscious mental episode, we are bound to admit that when we ascribe conscious thinking processes to persons, we are ascribing to them something which is different *toto coelo* from anything a computer scientist in sound mind would ascribe to his machines. 'The whole dignity of man', wrote Pascal, 'lies in his thought'; I do not believe that he shares this dignity, or some of its attendant problems, with Turing machines.

University of Singapore

Brit. J. Phil. Sci. 23 (1972), 13–23 *Printed in Great Britain*

Quantum-Theoretical Realism: Popper and Einstein v. Kochen and Specker[1]

by MICHAEL R. GARDNER

1 POPPER'S AND EINSTEIN'S REALISM

In his writings on quantum theory, early and late, Karl Popper has steadfastly maintained that the quantum-theoretic world-view is much closer to common sense than the theory's originators have claimed in their philosophical moments. As against Bohr, who held that quantum theory requires "a limitation of the well-defined application of space-time concepts" to atomic objects ([1966], p. 5) and consequently "a radical revision of our attitude towards the problem of physical reality" ([1961], p. 60), Popper has claimed that "the statistical laws of the theory . . . refer to a population of *particles* . . . which are . . . endowed with positions, and momenta (and mass-energy, and various other physical properties such as spin)" ([1967], p. 21; cf. [1935], Ch. IX). Presumably, then, his view is that for every system S and observable A, there is a real number r such that "$A = r$" is true of S.

Einstein held a view something like this, though it was somewhat more vague and speculative:

> For if the statistical quantum theory does not pretend to describe the individual system (and its development in time) completely, it appears unavoidable to look elsewhere for a complete description of the individual system. . . . Assuming the success of efforts to accomplish a complete description, the statistical quantum theory would, within the framework of future physics, take an approximately analogous position to the statistical mechanics within the framework of classical mechanics. I am rather firmly convinced that the development of theoretical physics will be of this type ([1949], p. 672).

2 KOCHEN AND SPECKER'S THEOREM

After several decades of controversy between the Bohr faction and the Popper–Einstein faction, a recent mathematical result in Kochen and Specker's [1967] is thought by some to have settled the issue in Bohr's

Received 22 March 1971

---

[1] Part of a doctoral dissertation, Harvard University, 1971. I am indebted to Hilary Putnam, Jeffrey Bub, and Sir Karl Popper for helpful criticisms of earlier drafts.

favour. Kochen and Specker begin with a precise statement of the programme which Einstein adumbrated in the above quotation. Suppose there is a "phase space" Ω and for each observable A a function f_A such that $f_A(w)$ is the value which a measurement of A would yield in a system characterised by the "hidden variable" $w \in \Omega$. Since one way to measure the observable $g(A)$ consists in measuring A and applying the function g to the result, for every w, $f_{g(A)}(w) = g(f_A(w))$, i.e.

$$(1) \quad f_{g(A)} = g(f_A).$$

A point in Ω constitutes the "complete description of the individual system" to which Einstein referred. The sense in which this "future physics" is to quantum theory as classical mechanics is to statistical mechanics is as follows. Our inability to predict the results of individual measurements arises solely from our ignorance of the values of the hidden variables. Dispersions in the results of measurements following a given preparation result solely from the physical impossibility of preparing an ensemble of systems each possessing the same value of w. If we knew the value of w for an individual system S, we could predict the result of any measurement upon it; and if we knew the probability distribution of S's locations in Ω, we could recover the quantum-theoretical statistics. Formally, there is associated with each quantum state ρ a probability measure μ_ρ on Ω such that $\mu_\rho(\Gamma)$ is the probability that w for a system in ρ is in the measureable subset Γ of Ω. Accordingly, for all observables A,

$$(2) \quad \mathrm{Tr}(\rho A) = \int_\Omega f_A(w) d\mu_\rho(w).$$

A space Ω, a set of functions f satisfying (1), and a set of probability measures μ_ρ satisfying (2) is called a *hidden variable theory.*

Kochen and Specker have shown that there is no such theory. We will sketch the proof. A set of observables (operators) is pairwise compatible (commuting) iff there is a single observable of which each is some function. If A_1 and A_2 are compatible and $A_1 = f_1(B)$ and $A_2 = f_2(B)$, we can define the sums and products

$$r_1A_1 + r_2A_2 = (r_1f_1 + r_2f_2)(B)$$
$$A_1A_2 = (f_1f_2)(B),$$

for all real r_1, r_2. The closure O of the set of observables under sums and products forms what Kochen and Specker call a *partial algebra* over the reals. It is easy to show (Bub [1969]) that condition (1) implies

$$(3) \quad f_{(k_1A_1+k_2A_2)} = k_1f_{A_1} + k_2f_{A_2}$$
$$(4) \quad f_{A_1A_2} = f_{A_1}f_{A_2}.$$

If 1 is the observable whose value is always the real number 1, and I the function whose value is always 1, then $f_1 = f_{I(A)} = I(f_A)$, i.e.

$$(5) \quad f_1 = I.$$

(3)–(5) are the defining conditions for a homomorphism f of the partial algebra of observables into the commutative algebra R^Ω, the set of all functions from Ω into the reals.

Let $Z_2 = \{0,1\}$. A partial algebra over Z_2 is called a *partial Boolean algebra*. The set of idempotents (elements such that $A^2 = A$) of a partial algebra forms a partial Boolean algebra, and the idempotents of a commutative algebra form a Boolean algebra. If A is an idempotent observable (i.e. a *question*), then

$$f_A = f_{A^2} = f_A f_A = (f_A)^2.$$

Hence, the restriction of f to the partial Boolean algebra Q of questions is a homomorphism into the Boolean algebra of idempotents of R^Ω. To each value of $w \epsilon \Omega$, there corresponds a map $h\colon O \to R$, namely $h(A) = f_A(w)$. Since

$$\begin{aligned} h(k_1A_1+k_2A_2) &= k_1h(A_1)+k_2h(A_2), \\ h(A_1A_2) &= h(A_1)h(A_2). \\ h(1) &= 1, \end{aligned}$$

h is a homomorphism. Its restriction to Q is a homomorphism onto Z_2. Accordingly, if q and p are compatible questions (commuting *projections*, or idempotent operators),

$$(6)\quad h(q+p) = h(q)+h(p)$$

$$(7)\quad h(qp) = h(q)h(p)$$

$$(8)\quad h(1) = 1.$$

Since $q \vee p = q+p-qp$ and $q \wedge p = qp$ (Jauch [1968], p. 38), (6)–(7) are jointly equivalent to

$$(9)\quad h(q \vee p) = h(q)+h(p)-h(p \wedge q)$$

$$(10)\quad h(q \wedge p) = h(q)h(p).$$

If the system's state-space is of dimensionality greater than two, one can easily show that no map satisfies (6)–(8) by using the method due to Bell ([1966]), who remarks (as do Kochen and Specker) that the result can also be obtained as a corollary of Gleason's theorem. (Gleason [1957]) It follows that there is no hidden variable theory.

In addition to their formal proof, Kochen and Specker give a direct intuitive argument against hidden variable theories. Consider a system in which

$$s_x^2+s_y^2+s_z^2 = 2\hbar^2,$$

and in which a measurement of the spin component along any axis always yields —1, 0, or 1 (letting $\hbar = 1$) Since the squares of the spin components in any three orthogonal directions are compatible, they may be measured simultaneously, and will always yield a zero and two ones. If there is a hidden variable theory, then there is a function h which correctly predicts the result of a measurement of the square of the spin component in each direction. These predictions may be indicated on a sphere by assigning a zero and two ones to every mutually orthogonal triple of points. By an elementary geometrical argument, Kochen and Specker show that such an assignment of numbers to points on a sphere is impossible.

3 FUTURE CONTINGENTS

At first this result may seem to be an antinomy. Any measurement of the square of a spin component will in fact yield either 1 or 0, and it must at least be possible to say truly, if not to know, which it will be. Or (Bub [1972]) was Aristotle right about tomorrow's sea battle after all? In the paper in which he first stated the result, Specker ([1960]) claimed that it shows that "no consistent predictions are possible concerning a quantum-mechanical system" whose Hilbert space is of dimensionality greater than two. He also suggested that his theorem is related to the Scholastic view that future contingents lack truth-values.

In fact, the Kochen and Specker results has no bearing upon these hoary issues. If we were to measure the squares of the components of spin in any three orthogonal directions x, y, and z, we would certainly obtain some value or other for each. Before the measurement it is possible to say, though perhaps not to know, what the results will be. It does not follow, however, that if we were instead to measure s_x^2, s_v^2, and s_w^2, where (x, v, w) is an orthogonal triple of directions different from (x, y, z), we would obtain the same result for s_x^2 we would have obtained in the first measurement. That the result of a measurement of one observable depends upon what other observables are measured with it is a consequence of Kochen and Specker's proof, and is clearly of interest outside of the context of hidden variable discussion. As Bell suggests ([1966]), it provides one way to understand Bohr's dictum ([1961], pp. 39 f.) concerning "the impossibility of any sharp distinction between the behaviour of atomic objects and the inter-action with the measuring instruments which serve to define the conditions under which the phenomena appear." But this dependence upon the experimental arrangement has no tendency to show that any quantum-mechanical predictions lack truth-values. "The measurement of s_x^2 I will perform here tomorrow at 3.00 PM will yield 1" has a truth-value. Kochen and Specker have shown only that what the truth-value is depends upon what observables I will measure simultaneously with s_x^2.

'Depends upon' does not mean 'is determined by.' It is an open question whether to each operator A and complete commuting set of operators including A, there is a unique number which A would yield if measured along with the rest of the set. An affirmative answer to this question is conjectured in the "hidden variable theory" of Bohm and Bub ([1966]). As Bub ([1969], [1972]) has made quite clear, because this theory does not assign a unique predicted value to each observable, it is not a hidden variable theory in Kochen and Specker's sense.

In one way, there is nothing mysterious or even specifically quantum-mechanical about this dependence of truth-values upon the total experimental arrangement. Whether it is true that I will eat lunch at noon tomorrow, after all, depends upon whether I will be hungry then. It certainly does not follow that the sentence 'I will eat lunch at noon tomorrow' lacks a truth-value. Yet in another way there is something strange and peculiarly quantum-mechanical about the specific dependence Kochen and Specker have shown to exist. What follows from their theorem is that if observables have values before they are measured, then measurements do not in general yield those values. (Otherwise, of course, the value yielded for one observable would not depend upon what other observables were measured simultaneously.)[1] But in the absence of any separate argument against the theory that measurements sometimes yield values different from those possessed before interaction with the measuring device, this corollary does not establish that observables sometimes lack values, and certainly does not establish that some propositions lack truth-values.

There may, however, seem to be another way to argue for certain truth-value gaps in quantum theory. Kochen and Specker have shown that there is no homomorphism of the partial Boolean algebra of questions onto $\{0,1\}$. Let us identify 1 with truth, 0 with falsity, and each question q with the "experimental proposition" which asserts that q has the value 1. A map h which assigns 1 or 0 to each question can then equally well be thought of as an assignment of a truth-value to each experimental proposition—i.e. proposition of the kind '$q = 1$'. However, since (9) and (10) resemble the usual conditions on assignments of truth-values to disjunctions and conjunctions,[2] one might think that in order to be an acceptable[3] assignment of truth-values to all the experimental propositions, h must satisfy (9) and (10), as well as (8). Accordingly, Bub [1972] has claimed that the non-existence of such a map implies that some experimental propositions lack truth-values.

---

[1] The Kochen and Specker result therefore quite clearly rules out the type of theory which Putnam espouses in his [1970]: "it is *true* that 'the particle has [e.g.] a momentum' (even if it is also true that 'the position is r_3'); and the momentum measurement merely finds this momentum . . .; it does not create, or disturb it in any way".

[2] A disjunction is true iff at least one disjunct is true, etc.

[3] 'Consistent' is Bub's term.

In arguing for these truth-value gaps, Bub assumes that any assignment of truth-values must give the ordinary meanings of 'or' and 'and' to 'v' and '∧' when they connect compatible propositions. However, the latter connectives are defined, not as truth-functions, but in terms of the operator algebra (§2, above). There is, therefore, no reason *a priori* to demand that assignments of truth-values accord with those of the standard calculus of truth-functions, as (9) and (10) require.[1] Of course, one might argue for the imposition of (9) and (10) on the ground that they follow from (1). Consideration of this suggestion will be deferred until the end of §4.

Bub has also suggested (*ibid.*) that Kochen and Specker's result shows that "a proposition is not true or false *simpliciter*, but has different truth-values relative to different appropriate [maximal] Boolean sub-algebras" of the questions, each of which corresponds to a complete commuting set of projections. Bub does not explain, however, how this relative notion of truth is to be defined. Certainly no relative term could satisfy Tarski's criterion of material adequacy for definitions of truth:

'———' is true iff ———,

which manifestly requires that 'true' be a one-place predicate of sentences. It is doubtful that this embodiment of our intuitive concept of truth could be modified in such a way as to allow truth to be relational. Accordingly, it is doubtful that any notion of truth as a relation between sentences and maximal Boolean subalgebras is intelligible. This is not to deny the intelligibility of hidden variable theories of the Bohm–Bub type, which predict different results of measuring an observable depending upon what observables are measured along with it, but only to deny that such theories imply that in quantum theory 'true' is a two-place rather than a one-place predicate. Whether I will eat lunch at noon tomorrow depends upon various things, but the truth of 'I will eat lunch at noon tomorrow' is not a relation between the sentence and those things.

4 RELEVANCE TO POPPER

There can be no doubt that Einstein's programme—recovering the quantum statistics from a distribution in an underlying phase space—is impossible, at least on Kochen and Specker's interpretation of the programme. Popper seems never to have asserted that such a recovery is possible. It is consistent to maintain—as he must—that every observable has a value and that there is some probability distribution of these values, but that one cannot recover the statistics of measurement results from this distribution because the distribution of unmeasured values differs from that of measured values. As we have already seen, a person who maintains that all unmeasured observables have values cannot maintain that measurements yield these values. So it is not surprising that he must also hold that the probabilities of these two types of values differ.

---

[1] The relation between 'v' and '∧' and the standard truth-functions is discussed in Gardner [1972*a*].

Though Popper need not accept Kochen and Specker's condition

$$(2) \quad Tr(\rho A) = \int_{\Omega} f_A(w) d\mu_\rho(w),$$

it would be a mistake to conclude for this reason that their result is irrelevant to his theory. Condition (2) is not used in their proof; only condition (1) is. And it would appear that Popper is at least committed to that. Suppose that associated with every A in O there is an x in R, the set of reals. Then let

$$\Omega = R^O = \{w | w\colon O \rightarrow R\}.$$

Define $f_A\colon \Omega \rightarrow R$ as follows:

$$f_A(w) = w(A).$$

If one of the functions w assigns the value x to A, it will presumably assign the value $g(x)$ to $g(A)$. We argued earlier that this is a necessary condition on assignments of measured values of observables, and it certainly seems at first sight a reasonable condition on unmeasured values as well. If $w(g(A)) = g(w(A))$, then $f_{g(A)}(w) = g(f_A(w))$; i.e. Popper's theory apparently has all of the features of a hidden variable theory used in Kochen and Specker's proof, and is therefore false.

While one can argue conclusively that assignments of measured values to the observables must satisfy condition (1), one can say in the case of unmeasured values only that it is plausible that they satisfy it, and also that it is rather difficult to see how values of A and $g(A)$ would be related if not by the function g. Surprisingly, however, condition (1) does have a consequence inconsistent with Popper's view, and should therefore perhaps not be attributed to him. Kochen and Specker show that (1) implies that to each $w \in \Omega$ there corresponds a homomorphism of the observables into the reals. In the case of the Ω defined above, h is just w. That w is a homomorphism implies that $w(A^2) = w^2(A)$. If A is a projection, then $w(A) = w^2(A)$. It follows that $w(A)$ is 1 or 0. This shows that if Popper accepts (1), he must also assign 1 or 0 to each question, measured or unmeasured.

But it is not at all reasonable to expect him to do this. Consider the spin example used by Kochen and Specker. Since s_z, whatever the direction of the z axis, yields -1, 0, or 1 (letting $\hbar = 1$) whenever measured, s_z^2 yields 1 or 0, and is therefore a question. Though Popper would predict 1 or 0 as the result of any measurement of s_z^2 (since it is a theorem that every measurement yields an eigenvalue), he would certainly not assign 1 or 0 as the (measured or unmeasured) square of the spin component in every direction. Such an assignment would be geometrically impossible. Suppose, for example, that for some fixed choice of the z axis, $s_z = 1$:

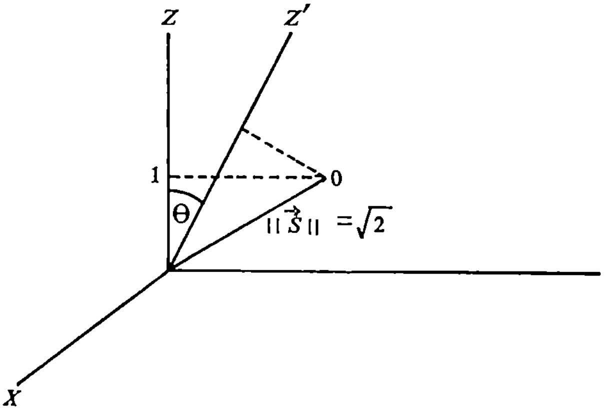


FIG. 1

Then, obviously, $\vec{s}$ will not simultaneously have a component of either —1, 0, or 1 along every other axis z', regardless of the orientation θ of z' with respect to z. Some of the unmeasured values of $s_{z'}$ will fall elsewhere in the interval [$-\sqrt{2}$, $\sqrt{2}$]. Accordingly, the unmeasured values of $s_{z'}^2$ may fall anywhere in the interval [0, 2].

The Kochen and Specker result does not, then, establish that Popper's view is inconsistent with the quantum theory. It does show, however, that his theory that every observable has a value commits him to holding that even though measurements of an observable A always yield members of its spectrum, A sometimes takes on other values, and also to holding that if two observables are functionally related, their values are not always related by the same function. It therefore seems impossible to say what the relation between the values of A and $g(A)$ is supposed to be. While this consequence may make Popper's theory somewhat unattractive —if only because of its lack of specificity—the previous paragraph suggests that the denial of condition (1) is not merely an *ad hoc* reply to Kochen and Specker but that the entire theory at least hangs together. Whether it thereby strangles is perhaps "a matter of personal belief", as Heisenberg might have said.

The essential reason why Kochen and Specker's condition (1) applies to measured values only is that the theorem is aimed at theories which seek to restore a certain kind of *determinism* in quantum theory. Specifically, its target is any theory which claims there is something (w) about a system S which determines (via the functions $f_A(w)$) what any measurement on S would yield.[1] Popper's concern, on the other hand, is not with determinism but ***realism***. He simply wishes to claim that it is at least possible (i.e., consistent with quantum theory) to maintain that unmeasured observables have values.[2] A realistic view need not be subject to the same strictures as an avowedly deterministic one.

Another way to state Popper's view, taking cognisance of Kochen and Specker, might run as follows. We must distinguish between (1) quantitative properties (or physical magnitudes), such as spin, position, etc., and (2) the corresponding observables which go under the same names. A value of an observable is by definition the result of a measurement of that observable or, equivalently, of the corresponding quantitative property. When a quantitative property is measured, its value always coincides with that of the corresponding observable. However, at other times, the quantitative properties may have values which they never have when measured because the algebraic structure of the observables prevents the corresponding observables from having those values. Hence, observables may not be conflated with quantitative properties of systems. The Hilbert-space operators correspond to observables, but not to quantitative properties. We should therefore not expect the algebraic relations among the observables (or the operators) to be mirrored in the algebraic relations among the values of the quantitative properties. (The distinction between observables and the corresponding quantitative properties, while useful here, will be ignored in the sequel for the sake of terminological orthodoxy.)

We can now take up the question left standing in section 3: whether the fact that (9) and (10) follow from (1) shows that they are reasonable conditions on any assignment of truth and falsity to all the experimental propositions. As we saw above, (1) is a reasonable condition upon a prediction (or retrodiction) of the results of measuring some observables, but not upon an assignment of simultaneous values to all observables, measured or unmeasured. Similarly, then, (9) and (10) are reasonable conditions upon a prediction of the results of testing some propositions, but not upon an assignment of simultaneous truth-values to all the propositions, tested or untested. Hence, the fact that no assignment of truth-values satisfies (8)–(10) does not show that no satisfactory assignment exists.

---

[1] Kochen and Specker do not require that the theories under attack be deterministic in the further sense that they specify the temporal evolution of w.

[2] Personal communication.

294092

5 CONCLUSION

We have seen that though it is easy enough to say that all observables have values, it is not only impossible to say precisely what these values are, but it is also impossible even to say what the relations are between the values of functionally related observables. Furthermore, Popper is forced to hold that the values of observables change when they are measured, but is presumably unable to say *how* they change. The theory therefore suffers from a severe lack of specificity. Perhaps partly for that reason, there does not seem to be anything conclusive to be said against it. On the other hand, it is not clear that there is anything conclusive to be said in its favour either. By comparison with other theories—e.g. Bohr's and Reichenbach's—as to which observables have values,[1] it has the advantage of not leading to famous paradoxes (Gardner [1972*b*]), and of being simple, in the sense of 'intuitively plausible.' But there is another version of quantum theory which is simpler than Popper's, in the sense of 'shorter'—namely, the minimal one which confines itself to statistical prediction of measurement results and takes no position at all on the question of when unmeasured observables have values. Since we only test the quantum theory by doing measurements, it is doubtful that any useful purpose is served by adopting *any* hypothesis—Bohr's, Reichenbach's, or Popper's—about the values, if any, of unmeasured observables. It would be a mistake, however, to go beyond this pragmatic consideration and argue that Popper's hypothesis is not merely superfluous and unspecific but inconsistent with the minimal statistical theory.

Harvard University

---

[1] Bohr held that statements about unmeasured observables are meaningless ([1961], pp. 25, 40), and Reichenbach held that they are neither true nor false ([1944], p. 145).

REFERENCES

BELL, J. S. [1966]: 'On the Problem of Hidden Variables in Quantum Mechanics', *Review of Modern Physics*, **38**, pp. 447–52.

BOHM, D. and BUB, J. [1966]: 'A Proposed Solution of the Measurement Problem in Quantum Mechanics by a Hidden Variable Theory', *Review of Modern Physics*, **38**, pp. 453–69.

BOHR, N. [1961]: *Atomic Physics and Human Knowledge.*

BOHR, N. [1966]: *Essays 1958–1962 on Atomic Physics and Human Knowledge.*

BUB, J. [1969]: 'What is a Hidden Variable Theory of Quantum Phenomena?', *International Journal of Theoretical Physics*, **2**, pp. 101–23.

BUB, J. [1972]: 'Hidden Variables and Quantum Logic', forthcoming in *Boston Studies in Philosophy of Science.*

EINSTEIN, A. [1949]: 'Reply to Criticisms'. P. A. Schilpp (*ed.*): In *Albert Einstein: Philosopher-Scientist*, 1949 pp. 665–88.

GARDNER, M. [1972*a*]: 'Is Quantum Logic Really Logic', forthcoming in *Philosophy of Science.*

GARDNER, M. [1972*b*]: 'Two Deviant Logics for Quantum Theory: Bohr and Reichenbach', forthcoming in *British Journal for the Philosopy of Science*, May 1972.

GLEASON, A. [1957]: 'Measures on Closed Subspaces of Hilbert Space', *Journal of Mathematics and Mechanics*, 6, pp. 885–93.

JAUCH, J. M. [1968]: *Foundations of Quantum Mechanics.*

KOCHEN, S. and SPECKER, E. P. [1967]: 'The Problem of Hidden Variable in Quantum Mechanics', *Journal of Mathematics and Mechanics*, 17, pp. 59–87.

POPPER, K. R. [1935]: *Die Logik der Forschung.*

POPPER, K. R. [1967] :'Quantum Mechanics without "The Observer"', in M. Bunge (*ed.*): *Quantum Theory and Reality*, 1967, pp. 7–44.

PUTNAM, H. [1970]: 'Is Logic Empirical?' in R. S. Cohen and M. Wartofsky (*eds.*): *Boston Studies in Philosophy of Science*, 5, pp. 216–41.

REICHENBACH, H. [1944]: *Philosophical Foundations of Quantum Mechanics.*

SPECKER, E. P. [1960]: 'Die Logik nicht gleichzeitig entscheidbarer Aussagen', *Dialectica*, 14, pp. 239–46.

Empirical Simplicity as Testability*

by KENNETH S. FRIEDMAN

Given the nature of the role played by our intuitions of empirical simplicity in scientific investigation, the development of formal criteria of empirical simplicity is an important task for the philosophy of science. So far, however, attempts to give such formal criteria have not been successful. The purpose of this paper is first to criticise three unsuccessful criteria of simplicity, and second, to propose an alternative criterion.

The most straightforward approach to empirical simplicity links the simplicity of a hypothesis to its logical strength. Along these lines one might maintain that the logically weaker of two hypothesis (which is also the more probable and the more verifiable of the two) is the simpler one. This leads to unintuitive results. To borrow the example of Nelson Goodman ([1961]), "All maples, except perhaps those in Eagleville, are deciduous" is weaker, but not as simple as "All maples are deciduous". Moreover, if our aim in science is to tentatively adopt the simplest hypothesis consistent with the facts, we are bound to be frustrated; there is no such thing. For *any* hypothesis consistent with the facts, there is a weaker, hence simpler one also consistent with the facts. Besides, science strives more for strong than for safe hypotheses.

A diametrically opposed criterion—that the stronger, and so less probable and more falsifiable of two hypotheses is the simpler one—has been suggested by Karl Popper in *The Logic of Scientific Discovery*. This, too, leads to problems. For given any hypothesis, one can construct a stronger, hence simpler one just by adding a non-trivial conjunct. Also, to borrow again from Nelson Goodman, "All maples whatsoever, and all sassafras trees in Eagleville, are deciduous" is stronger, but not as simple as "All maples are deciduous".[1]

A third criterion, also suggested by Popper, deserves closer attention. This criterion, called a theory of dimension, is based on the intuition that straight lines are simpler than circles which in turn are simpler than parabolas, because it takes fewer points (or parameters) to determine a straight line than to determine a circle, and fewer to determine a circle

Received 1 *October* 1970

---

* I am grateful to the referees of *The British Journal for the Philosophy of Science* for their helpful comments on an earlier draft of this paper.

[1] This criterion has also been criticised by Barker ([1961]).

than to determine a parabola. To oversimplify a little, the number of points required to determine a figure is the dimension of that figure; the smaller the dimension, the simpler the figure.

Popper writes (Popper [1935], pp. 285–6):

> The proposed definition says: A theory t is called d-dimensional with respect to the field of application F if and only if the following relation holds between t and F: there is a number d such that (a) the theory does not clash with any d-tuple of that field and (b) any given d-tuple in conjunction with the theory divides all the remaining relatively atomic statements uniquely into two infinite subclasses A and B, such that the following conditions are satisfied: (i) every statement of the class A forms, when conjoined with the given d-tuple, a "falsifying $d+1$-tuple" i.e. a potential falsifier of the theory; (ii) the class B on the other hand is the sum of classes [B_i] such that the conjunction of any number of statements belonging to any one of these subclasses [B_i] is compatible with the conjunction of the d-tuple and the theory.

As an example, the hypothesis "All borks move in straight lines" has a dimension of 2, because that hypothesis cannot clash with any pair of statements of the form "This bork is at point p". Moreover, given that this bork passes through points p_1 and p_2, all other points belong to one of two classes. Points not collinear with p_1 and p_2 belong to class A, for if the bork passes through any of these points, having passed through p_1 and p_2, the hypothesis is falsified. Points collinear with p_1 and p_2 belong to class B. The bork can pass through any number of them without falsifying the hypothesis. By similar reasoning, the hypothesis "All borks move in circles" has a dimension of 3, and so is not as simple as "All borks move in straight lines".

While this criterion seems to agree with our intuitions, at least in most cases to which it applies, there are serious problems to consider. One problem arises from the ways in which the dimension of a theory can be reduced. We can reduce the dimension of the hypothesis H "All borks move in ellipses" by specifying a point p on the ellipse to get the new hypothesis H' "All borks move in ellipses which pass through the point p". However, we have not simplified our original hypothesis; H is as simple as H'. We can also reduce the dimension by equating the eccentricity of the ellipse to 0, getting "All borks move in circles". In this case we have simplified our original hypothesis.

It is obvious that more is at work here than just the dimension of a theory. Popper tries to take this into account by distinguishing between a material reduction and a formal reduction of dimension. A material reduction introduces proper names (or place names) into the hypothesis; a formal reduction does not. Thus the former reduction, specifying point p, is a material reduction, and the latter is a formal reduction. Popper seems to maintain that all and only formal (as opposed to material) reductions of dimension are relevant to simplicity (p. 385). However, consider the hypothesis *H''* "All borks move in ellipses with eccentricity ·53". This is a formal reduction of dimension, since it mentions neither place nor proper names. Yet it is not clear that any intuitions of simplicity conform to the view that *H''*, but not *H'*, is simpler than *H*.

There is a second problem for this criterion. Consider the following configuration of points · · : No purely convex figure—circle, ellipse, parabola, convex polygon—can pass through all of these points. Thus on Popper's criterion, the dimensions of these figures can be no greater than 3. But through any three points which are not collinear an uncountable number of ellipses, parabolas, squares, etc., can be drawn, so the dimensions of these figures cannot be smaller than 4. These conclusions imply that ellipses, parabolas, convex polygons, cannot be assigned any dimension.

I will now present my own criterion for comparing the empirical simplicity of two hypotheses. According to this criterion the more testable a hypothesis is, the simpler it is. Note, however, that this involves a non-standard version of testability—not equivalent to falsifiability (although it has clearly been influenced by Popper's contributions). In particular, the conjunction of any two empirical statements is always more falsifiable (hence, according to the standard version, simpler) than either conjunct. By contrast, according to this non-standard version, in some cases a conjunction may be less simple than its conjuncts.

Perhaps an illustration will make the motivation for this different version of testability clearer. Consider an empirical hypothesis of the form $(x)(Ax \rightarrow Bx)$. The potential falsifiers of this hypothesis are of the form $Aa \ \& -Ba$. One can falsify the hypothesis by, among other things, establishing the truth of *Aa* and the falsity of *Ba*. One can corroborate the hypothesis by, among other things, establishing the truth of *Aa* and also of *Ba*. A test is a procedure which either falsifies or corroborates a hypothesis, relative, of course, to background knowledge. (It does not just falsify the hypothesis, for hypotheses do survive tests and are sometimes corroborated by them.)

On one hand, the falsifiability of a theory is proportional to the logical ease of falsifying it, which can be expressed in terms of the amount of information necessary to establish that a potential falsifier is true. On the other hand, the testability of a theory (on this new non-standard version) is proportional to the logical ease of testing it, which can be expressed in terms of the amount of information necessary to determine whether a potential falsifier is true or false.[1] Establishing criteria for falsifiability is straightforward. The stronger the theory, the weaker its potential falsifiers; the weaker its potential falsifiers, the less information required to establish their truth. Establishing criteria for testability, however, is not so straightforward, and the rest of this paper will be concerned with developing these criteria.

The first step in this procedure will be to determine criteria for comparing the testability (and empirical simplicity) of two predicates. This is based on the notion that if a test for one predicate (i.e. a test which determines whether that predicate applies to a given object) depends on certain information above and beyond that required in a test for a second predicate, then the second predicate is more testable and simpler than the first. The problem is to spell out the nature of this information and also the nature of the dependence of a test on this information. Intuitively, this information does not by itself determine whether or not one predicate applies, but in conjunction with whether one predicate applies does determine whether the other one applies. Further elaboration requires the clarification of some preliminary points.

A test for a predicate is a procedure which in appropriate circumstances, either determines (relative to background knowledge) that the predicate applies to an object in question or determines that it does not apply. Thus looking at an object through a red filter is a test for its being green, for if light is transmitted through the filter, the object cannot be green. A test for one predicate is independent of a second predicate if all its possible results are compatible with the applicability of the second predicate, and also compatible with the inapplicability of that predicate to the test object. For example, a test for green by observing an object through a red filter is independent of the predicate 'is square', for light being transmitted and light not being transmitted are each compatible with the object's being square and also with the object's not being square. However, such a test for green is not independent of the predicate 'is blue', since light being transmitted is not compatible with the object's being blue.

Suppose predicate *Q* 'contains' the predicate *P*, i.e. once we know whether *P* applies, the application of *Q* is solely dependent on the application of some predicate *I*. We say that *I* is information for *Q* dependent on *P*

---

[1] This approach, based on the notion of the amount of information necessary to test a theory, (though arrived at independently) is similar to the one adopted by Giedymin ([1964]). However, Giedymin's approach is based on the amount of information necessary to falsify a theory, whereas this paper is based on the amount of information necessary to test (not just falsify) a theory. The purpose of this paper, moreover, to provide a formal, unified account of empirical simplicity, differs radically from the purpose of Giedymin's paper.

(in that if it is given that P applies—or given that it does not—I constitutes sufficient information to determine whether or not Q applies). If I is information for Q dependent on P and is not information for Q dependent on any other predicate, then I is said to be information for Q parasitic on P (in that it constitutes information sufficient to determine the applicability of Q, but if and only if it is given that P applies—or given that it does not). In other words, I is information for Q parasitic on P if and only if given P (or $-P$), but not any other predicate in place of P, I (or $-I$) is equivalent to Q.

Formal criteria for all such parasitic information are: a predicate I constitutes information for Q parasitic on P if and only if it satisfies either the first two of the following conditions or the third condition.

1. $(x)(Px \mathbin{\&} Ix \,.\rightarrow Qx : \& : Px \mathbin{\&} -Ix \,.\rightarrow -Qx)$ (or $P/-P$, $Q/-Q$, $I/-I$)
2. $-(\exists R)(x)[(Rx \mathbin{\&} Ix \,.\rightarrow Qx : \& : Rx \mathbin{\&} -Ix \,.\rightarrow -Qx) \mathbin{\&} -(Rx \leftrightarrow Px)]$ (or $P/-P$, $Q/-Q$, $I/-I$)
3. $(x)(I^{*}x \rightarrow Ix \,.\, v \,.\, Ix \rightarrow I^{*}x)$ (or $I/-I$, I^{*} satisfies 1 and 2)

We denote by 'test $(Q{:}I_P)$' any test for Q that is independent of all I such that I is information for Q parasitic on P. We designate $T(Q{:}I_P)$ as the set whose elements are tests $(Q{:}I_P)$. PREDICATE P IS SIMPLER THAN PREDICATE Q IF AND ONLY IF $T(Q{:}I_P) \subset T(P{:}I_Q)$. Several examples may clarify both the nature and the implications of this formalism.

Ex. 1. P—is square Q—is green and square

Suppose I stands for 'is green'. Condition 1 is met, for anything that is square and that is green is green and square, and anything that is square and that is not green is not green and square. Condition 2 is met, for given the meanings of I and Q, no predicate other than 'is square' satisfies the first condition. Thus I is information for Q parasitic on P, and a test for Q is a test $(Q{:}I_P)$ only if it is independent of 'is green'. By establishing that P applies we need not establish whether Q applies. So establishing that P applies is a test $(P{:}I_Q)$, but not a test $(Q{:}I_P)$. Any test that establishes that Q applies *a fortiori* establishes that P applies. Any test that establishes that Q does not apply but does not establish that P does not must establish that the object in question is not green. Such a test would not be independent of 'is green', and so would not be a test $(Q{:}I_P)$. Then $T(Q{:}I_P) \subset T(P{:}I_Q)$, and so P is simpler than Q.

This seems in better agreement with our intuitions than the corresponding result obtained from Popper's criterion, that Q is simpler than P.

---

[1] $P/-P$ means 'P uniformly replaced by $-P$'.

Ex. 2. P—is square. Q—is green or square

The reasoning here is similar to that of the first example. The only difference is that in the first condition one must substitute $-P$ for P. The first condition is now satisfied because anything that is not square and that is green is either green or square, and anything that is not square and is not green is not either green or square. The second condition is satisfied as above. Thus a test $(Q:I_P)$ must be independent of 'is green'. By establishing that P does not apply we do not necessarily establish whether Q applies. So establishing that P does not apply is a test $(P:I_Q)$, but not a test $(Q:I_P)$. Any test showing that Q does not apply *a fortiori* shows that P does not apply. Any test showing that Q applies which does not show that P applies must establish that the object in question is green. Then this cannot be a test $(Q:I_P)$. So $T(Q:I_P) \subset T(P:I_Q)$, and P is simpler than Q.

These two examples show that testability and empirical simplicity are not determined by the relative logical strength of predicates. In the first example the weaker predicate is the more testable (and simpler) one; in the second example the stronger is the more testable (and simpler) one.

Ex. 3. P—is rectangular. Q—is rectangular and equilateral, i.e. is square.

In this example a test $(Q:I_P)$ need not be independent of 'is equilateral'. This is because of condition 2, in which we might let R be 'has two diagonals, which are equal to each other'. Determining that P applies does not necessarily determine that Q applies and so furnishes a test $(P:I_Q)$ which is not a test $(Q:I_P)$. Showing that the object in question is not equilateral shows that Q does not apply, but not that P does not. This, then, constitutes a test $(Q:I_P)$ which is not a test $(P:I_Q)$. So $T(Q:I_P)\ \boldsymbol{e}\ T(P:I_Q)$, and Q is as simple as P.

The point of this example is to show that simplicity is not determined by logical form. For P has the same logical form in the first and third examples, and so does Q; yet in the first example P is simpler than Q and in the third it is not.

Ex. 4. P—is rectangular. Q—is square or green.

As in the previous example, showing that P applies does not show whether Q applies and so furnishes a test $(P:I_Q)$ that is not a test $(Q:I_P)$. Finding that the object is green shows that Q does apply, but not whether P applies. This, then, constitutes a test $(Q:I_P)$ that is not a test $(P:I_Q)$. So $T(Q:I_P) \not\subset T(P:I_Q)$, and so Q is as simple as P.

This example indicates a problem for my criterion of simplicity, for the relation 'as simple as' is not transitive. 'Is square or green' is as simple as 'is rectangular', and 'is rectangular' is as simple as 'is square', but 'is square or green' is not as simple as 'is square'. This sort of result can be avoided by amending our definition of simplicity.

I will first define a composite predicate as one which is logically equivalent to some concatenation of other predicates such that, according to the present criterion of simplicity, these other predicates are simpler than the composite one. (Thus, 'is square or green' is a composite predicate, but 'is rectangular and equilateral' is not.) Predicates which are not composite are simple.[1] Two simple predicates, P and P', are similar if and only if P $(-P)$ is incompatible with P' $(-P')$. Two composite predicates are similar if and only if when they are expressed in simplest terms of their simple predicates, there is an isomorphism from each simple predicate of one onto a similar predicate of the other.

We define a descending chain from P_i to P_j as a sequence of predicates such that (1) at least one predicate is simpler than its successor, and (2) each predicate is either simpler than or similar to its successor. We define a chain from P_i to P_j as a sequence of predicates, each one similar to its successor.

We now say that P_i is simpler than P_j if there is a descending chain from P_i to P_j; if P_i is simpler than P_j, P_j is less simple than P_i. P_i is as simple as P_j if there is a chain from P_i to P_j. P_i and P_j are incomparable if there is neither chain nor descending chain between them.

Several interesting results follow from these definitions. 'Is rectangular' is now simpler than 'is square or green', for 'is rectangular' is similar to 'is square' which in turn is simpler than 'is square or green'.[2] We have some general results as well. For any two predicates, P_i and P_j, P_i is either simpler than, as simple as, less simple than, or incomparable with P_j. If P_i is simpler than P_j and P_j is simpler than or as simple as P_k, then P_i is simpler than P_k. If P_i is as simple as P_j and P_j is as simple as P_k, then P_i is as simple as P_k. If P_i is either simpler than, as simple as, or less simple than P_j and P_j is incomparable with P_k, then P_i is incomparable with P_k.[3]

It is a simple matter to extend this criterion for simplicity of predicates to hypotheses. Hypothesis H1 is simpler than hypothesis H2 provided (1) H2 has at least as many predicates as H1, and (2) the predicates of H1 and H2 can be put into correspondence such that (*a*) no predicate of H1 corresponds to a simpler predicate of H2, and (*b*) if H1 has as many predicates as H2, some predicate of H1 is simpler than a corresponding predicate of H2.

---

[1] This statement is not taken to imply a commitment to absolutely simple predicates. Whether a predicate is simple or composite depends on language, background beliefs, and interests.

[2] This is interesting because 'is rectangular' and 'is square or green' have incomparable logical strengths, and so are incomparable on Popper's theory of simplicity.

[3] This theory of simplicity also provides an answer to Goodman's new riddle of induction. It is possible to show, relative to certain weak assumptions, that the predicate 'green' is simpler than the predicate 'grue'.

One can extend this criterion of simplicity to account for the intuitions discussed in the theory of dimension. This can be done by modifying our criteria for tests. First, since we are testing geometrical figures by observing individual points, it is important to note that while a finite number of points can entail that a geometrical predicate does not apply (e.g. that a particular figure is not a square), it takes an uncountable number of points to entail that it is a square. In view of this it is reasonable to weaken our requirements for positive tests (or confirmations). I will say that the applicability of a predicate is confirmed by a test if, provided that it had a finite prior probability of applying, its posterior probability of applying, given the test, equals 1.

Second, since space-time points do not figure in the identity conditions of experiments, I will define a test* for a predicate such that tests which differ only in specifying different space-time points count as the same test*. Thus all n-tuples (i.e. statements of the form "Points $p_1 \ldots p_n$ lie on the orbit of the object in question") with the same n constitute the same test*. In particular, a test* for a predicate P is the set of all n-tuples of every n such that every n-tuple of that n constitutes a test for P. For example, a test* for 'is a straight line' is the set of all n-tuples, $n>2$. For given any two points a straight line is determined, and any other points comprising an n-tuple either (*a*) all lie on that line, in which case we will see that the n-tuple confirms the applicability of 'is a straight line', or (*b*) some do not lie on that line, in which case the n-tuple shows the inapplicability of 'is a straight line'.

This suggests the following definition of the dimension of a predicate. The dimension of a predicate P is the number n of the smallest n-tuple of basic statements which divides all other basic statements s (of the form "Point s lies on the orbit of the object in question") into two mutually exclusive classes[1]:

C1 $\{\hat{s}/Px \mathbin{\&} n\text{-tuple} \,.\rightarrow s\}$

C2 $\{\hat{s}/s \mathbin{\&} n\text{-tuple} \,.\rightarrow -Px\}$

Let us see how this criterion operates. Suppose we are wondering whether the predicate 'is a circle' applies to a particular orbit. Given three points on the orbit, we have no way of knowing whether that predicate applies. Now add a fourth point. Either this lies on the circle determined by the first three points or it does not. If it does not, the basic statement corresponding to that point is a member of class C2 as it shows that the predicate does not apply. If it does, the conjunction of the predicate and the 3-tuple implies that that point lies on the orbit, and so the basic statement is a member of class C1. Thus 3 points divide the rest of space into the two classes C1 and C2, and consequently, the dimension of the predicate 'is a circle' is 3.

---

[1] As in n. 1, p. 31, this is not taken to imply a commitment to absolutely basic statements, independent of background knowledge, theoretical structure, etc.

It remains to be shown that any member of C1 confirms the applicability of predicate P. Writing p as an abbreviation for "The predicate P applies to the entity in question", we use Bayes's theorem to write

$$P(p/s \in \mathrm{C1}) = \frac{P(s \in \mathrm{C1}/p)P(p)}{P(s \in \mathrm{C1}/p)P(p) + P(s \in \mathrm{C1}/-p)P(-p)}$$

For predicates one generally considers in this context, e.g. 'is a sphere' or 'is sinusoidal', the Lebesgue measure of the orbit relative to the measure of the space in which it occurs is 0. (In the above example the Lebesgue measure of the circumference of a circle in two- or three-dimensional space is zero.) If we take this as a probability measure, then $P(s \in \mathrm{C1}/-p) = 0$. So if $P(p) > 0$, then $P(p/s \in \mathrm{C1}) = 1$. Thus the applicability of the predicate is confirmed.

Given our definition of dimension, any n-tuple constitutes a test* for a predicate of dimension $n-1$ or smaller, though not a test* for a predicate of dimension n or greater. Then the set of tests* for a predicate of dimension n is a proper subset of the set of tests* for a predicate of any smaller dimension. Given the result of substituting 'test*' for 'test' in our definition of simplicity—PREDICATE P IS SIMPLER THAN PREDICATE Q IF AND ONLY IF $T^*(Q{:}I_P) \subset T^*(P{:}I_Q)$—a predicate of given dimension is simpler than any predicate of greater dimension.

What I have tried to do here is to provide an alternative theory of simplicity to Popper's such that (1) this theory, like Popper's, explains the methodological significance of simplicity, and (2) this theory is in closer agreement with our intuitive notions of simplicity and is immune to certain counter-examples brought against Popper's theory.

State University of New York at Oswego

REFERENCES

BARKER, S. [1961]: 'On Simplicity in Empirical Theories', *Philosophy of Science*, **28**, pp. 162–71.

GIEDYMIN, J. [1964]: 'Strength, Confirmation, Compatibility', in M. Bunge (*ed.*): *The Critical Approach to Science and Philosophy*, 1964, pp. 52–60.

GOODMAN, N. [1961]: 'Safety, Strength, Simplicity', *Philosophy of Science*, **28**, pp. 150–1.

POPPER, K. [1935]: *Die Logik der Forschung*. (Page references refer to the 1959 English translation).

Brit. J. Phil. Sci. 23 (1972), 35–43 *Printed in Great Britain*

The Propensity Theory of Probability

by ALAN R. WHITE

A suggestion about *probability* made by C. S. Peirce (cf. his [1910]) has lately found favour among contemporary philosophers of science under the title of 'The Propensity Theory of Probability'.[1] According to Peirce, 'The statement [*sc.* the probability that if a die be thrown from a dice box it will turn up a number divisible by three is one-third] means that the die has a certain "would-be" . . . analogous to any *habit* that a man might have'. Popper's recent resuscitation of the theory expresses it thus: 'Every experimental arrangement is *liable to produce* . . . a sequence with frequencies . . . [which] may be called probabilities . . . [which] may be looked upon as *properties of this arrangement. They characterise the disposition, or the propensity,* of the experimental arrangement to give rise to certain characteristic frequencies, *when the experiment is often repeated*' (Popper [1957], p. 67).

Though I sympathise with the desire of these philosophers to give a non-subjective analysis of probability—which I believe can be done—I shall argue that this particular theory is mistaken; that, in fact, what it does is to assimilate the idea of *being likely* to the idea of *being liable.*

An initial difficulty for any examination of the theory is that its proponents differ on five important points. (1) Is the theory intended to give an *analysis* of the notion of probability, as Peirce suggested, or to provide a *hypothesis* about the nature of the world, as Popper contends? (2) Is the notion of a *propensity* to be interpreted in terms of a *habit,* as Peirce took it, or in terms of a *liability,* as Popper takes it, or as a mixture of both? (3) What is it that is alleged to have the propensity? Is it a physical object, such as a die or a coin, as Peirce originally held, or is it an 'experimental arrangement' or a 'set up' of, e.g. dice or coin tossing, as Popper now suggests? (4) What is the propensity by which the notion of probability is analysed or interpreted? Is it such a property as 'true', 'loaded', 'unbiased', etc., or is it the propensity to have a certain probability? (5) Is

Received 22 July 1971

---

[1] Cf. e.g. Popper [1957], [1959] and [1967], esp. § 8; Hacking [1965]; Mellor [1969] and Levi [1967], ch. 14.

the theory intended to be applicable only to statistical examples, e.g. of coin tossing or atomic radiation, or to all cases of probability.

Let us consider these five points in turn.

(1) There is no doubt that most proponents of the theory have thought of it as an *analysis* of at least one notion of probability. Peirce explicitly said that he was trying 'to define the meanings of the statement that the probability that . . . is . . .'. Hacking says his task is 'to define a property', which he first calls 'long run frequency' and later 'chance'. Mellor states his aim as 'to analyse statistical probability, which I call "chance"'. Levi, following Popper, usually calls it 'a propensity interpretation' or 'a propensity account' of probability; but he also speaks of statistical probabilities as being dispositions and considers a statement of probability as 'truth functionally equivalent' to its corresponding dispositional statement.

Popper, on the other hand, in calling his view the 'propensity interpretation' of probability, explicitly disavows that he is giving an analysis of the meaning of the word 'probability'. He says that what he gives is 'the physical (or perhaps metaphysical) hypothesis . . . that every experimental arrangement . . . generates physical propensities which can be tested by frequencies'. Yet he often speaks of his interpretation as a rival to the 'subjective and objective interpretations' which have usually been considered as analyses of the notion of probability or as theories of probability; he takes interpretation to be the interpretation of a statement and to signify what the statement is 'interpreted to mean'. Furthermore he explicitly asserts that 'probabilities are physical propensities'.

It seems plausible, therefore, to consider this theory as a philosophical analysis of the notion of probability.

(2) What do the authors of the theory mean by a 'propensity'? Peirce, who spoke of the die's 'would-be' rather than of any 'propensity', explicitly compared this to a human habit. Indeed, he sometimes speaks of the die's habit. Modern exponents, using a variety of words, such as 'propensity', 'disposition', 'tendency', 'urge', 'liability', have compared the probability of something's happening to such properties as the solubility of salt, the brittleness of a glass, Newtonian forces, the inertial mass of an object and the generosity of a man.

There is a vital difference between a habit or tendency on the one hand and a liability or predisposition on the other which is obscured by such a general notion as disposition or propensity. This is the difference between 'what does happen when' and 'what would happen if', between what is of regular or invariable occurrence and what is of potential occurrence.

One could not say of a man who had never lost any papers or of a clock which had never lost five minutes a day that the man had a habit of losing his papers or that the clock had a tendency to lose time. But one could properly say of a man who is generally absent minded that he is liable to lose any papers entrusted to him and of a clock with a very rough movement that it is liable to lose or gain time, although the man has not yet been given any papers or the clock yet been wound up. A piece of brittle glass need never have broken nor a pinch of soluble salt ever have dissolved.

This important difference between what has a tendency to V or is in the habit of Ving and what is liable or predisposed to V is obscured because something's having a tendency to V or being in the habit of Ving is often a good reason for supposing it to be liable or predisposed to V. A car which has a tendency to over-steer is more liable to over-steer on the next corner than a car which has exhibited no such tendency. But it is a mistake to assimilate the tendency to the liability. It is, indeed, the same sort of mistake as that which confuses the frequency with which something has happened with either its liability to happen again or—what we shall see is different—the probability of its happening again. A kindred mistake is this. The fact that a type or a collection of things has a tendency to V, e.g. students have a tendency to become nervous before examinations, may provide good grounds for the supposition that an individual member of this collection or an instance of this sort is liable to V, but not that it has a tendency to V, since it may never have Ved. The student who has never before approached an examination can be liable to, but not tend to, be nervous. The unused car in the showroom cannot have acquired any tendencies. That radium atoms tend to V may show that an individual atom is liable to, but not that it tends to, V.

In one respect a tendency or a habit is a more plausible candidate than a liability or predisposition for the role of probability, since the fact that something has commonly or normally Ved is a better reason for supposing that it will V—and, *a fortiori*, for supposing that it is probable that it will V—than the fact that it is liable or predisposed to V. It is clear, however, that the probability of something's Ving cannot be the same as its habit of Ving since it need not have any such habit. Indeed, nothing can have the habit of doing a particular thing for the first time, but it can be probable that it will do it for the first time.

Since, therefore, it is clearly mistaken to suppose that the probability of so-and-so's being such-and-such can be the same as its habit, if any, of being such-and-such and since Peirce himself also spoke of the 'would-be' of a die or a coin, the only plausible interpretation of the propensity theory is in terms of a liability and the only plausible analogy for probability is a property like solubility or brittleness. Moreover, despite their use of such general words as 'disposition', this seems to be the modern version of the theory.[1] It is, therefore, the one I shall henceforth examine.

(3) To what does the propensity theory attribute this propensity or liability? First, Peirce certainly thought it was such a thing as a die or a coin just as it is a man who has a habit. In this he is followed by several contemporary writers. For example, Cooper, who thinks Peirce, Von Mises and Popper all agree with him, subscribes, in his [1965], to the view that 'the probability of an object's undergoing a change is a property of the object, a "would be" or a propensity'. Mellor, who, however, realises that Popper and Hacking differ from him, also attributes the propensity to an object such as a coin or a die. Levi gives the fairness of a coin as one example of the property. According to him, 'A statistical hypothesis that assigns a statistical probability distribution to outcomes of a draw from the urn in effect attributes a certain statistical disposition to the urn'. Secondly, it is usually, perhaps always, such things as these which can properly be said to have a propensity or to be liable to V. Thirdly, such a subject of the propensity would best suit this theory's comparison of probability to such dispositional properties as solubility, brittleness, force, etc., which are properties of similar objects.

Unfortunately for the theory, as thus interpreted, probability, unlike a propensity, liability or habit, is not in fact attributed to such things as coins, dice, people, weather, etc. It is not the probability of the coin, but of the coin's falling heads, that is in question; not the probability of me, but of my living to be 80; not the probability of Tennyson, but of his having written such-and-such a poem. Such phrases as 'the probability of heads, of rain, of an early death, of a change of government' can mislead, since what these signify is the probability of there being, or that there will be, heads, rain, an early death or a change of government. In exactly the same way *chance* is something which is attributable to, e.g. the coin's falling heads or my living to be 80 and not to the coin or to me. Likewise, the chances of heads or of a change of government are the chances of there being heads or a change of government. Similarly, though something may be both liable and likely to V, it has a liability, but not a likelihood, to V. The only likelihood there is—which is not a likelihood it has—is of its Ving.

A propensity, liability or habit, on the other hand, is not something which is attributable to a coin's falling heads, but to a coin; not something attributable to my losing my temper, but to me. Contrariwise, where a probability is attributable to a thing, such as heads, rain, an early death or a change of government, it would be nonsense to suggest that any of these could have a propensity. Heads, rain or a change of government may be likely, but they cannot be liable.

---

[1] E.g. Hacking and Mellor admit that a disposition need never manifest itself. (Cf. Hacking [1965], p. 10, and Mellor [1969], pp. 24, 27.)

In short, that to which can be attributed a propensity, liability or habit, like that which can be soluble or brittle, unbiased, fair, true or loaded, is not the sort of thing to which can be attributed a probability or likelihood; nor is that to which can be attributed a probability the sort of thing to which can be attributed a propensity. It is because *probability* and *propensity* are categorially different that the sorts of things to which they can be attributed are categorially different. Once again what the theory has confused is the, possibly correct, supposition that the propensity of something to V, e.g. of a coin to fall heads and tails in equal proportions, makes it probable that it will V with the fallacious theory that its propensity to V is the same as the probability of its Ving. It has misinterpreted the relation of 'liable to V' and 'likely to V' as an equivalence.

Contemporary attempts to substitute some other subject as the bearer of that propensity which probability is alleged to be run into similar difficulties. Popper, for instance, usually[1] attributes the propensity to what he calls 'the experimental arrangement' or—what he says is the same—'the generating conditions', in accordance with which an event, such as obtaining a 6 on the next throw, is produced. Since, however, he sometimes equates both the experimental arrangements and the generating conditions with 'the singular event', it looks as if he is really attributing the propensity to an event. But, as we have seen, though events can have a probability, they cannot have a propensity and though experimental arrangements can, perhaps, have a propensity, they cannot have a probability; whereas dice or pennies which can have a propensity, but not a probability, are said by Popper not to be things which, on his theory, have either the probability or the propensity. Part of the difficulty in discovering to what exactly Popper does attribute either a probability or a propensity is due to his tendency to argue that to be a property of something is, or is a consequence of, being dependent on it. Thus, because the probability of something depends on certain experimental conditions, the probability is often attributed to, or said to be a property of, these conditions.

Hacking holds that the propensity belongs to what he calls 'a chance set-up', e.g. a coin together with a tossing device, a stretch of road and those who drive on it, a piece of radium together with a recording mechanism, even a pair of mice. But to none of these, any more than to the coin, the road or the radium itself, can a chance or a probability sensibly be attributed.

---

[1] E.g. in his [1957], p. 60; [1959], pp. 31, 35, 37 and [1967], pp. 32–3 and 38–40. But, cf. his [1967], p. 32, where he speaks of 'the propensity of a possibility to realise itself upon repetition'.

Mellor also speaks of 'a chance set-up', but he differs from Popper and Hacking in that, first, he follows Peirce in equating this chance set-up which has a propensity with such a thing as a coin or atom of radium[1] and not with the combination of any of these and the experimental conditions. Secondly, Mellor does not equate chance or probability with the propensity itself, but with its display. Hence, he attributes a propensity to what can properly have one, e.g. a coin, and a probability to what can properly have one, e.g. a coin's falling heads.[2] In other words, he makes *likely to be* a display of *liable to be* and not equivalent to it. His mistake is to suppose that probability must be the display of a propensity; that *likely to be* is possible only as a display of *liable to be.*

(4) This brings us to the related question: What is the propensity by which the notion of probability is analysed or interpreted?

If it is the propensity to have such-and-such a probability or chance—or, as Mellor often says, a chance distribution—then this account seems to have no more virtue than that of the notorious 'dormitive virtue' to explain falling asleep. Thus Peirce defines the die's 'habit' or 'would-be' in terms of how it would behave if it underwent an endless series of throws from a dice box. Levi supposes that assigning a statistical probability distribution to outcomes of draws from an urn is attributing a statistical disposition to the urn. Hacking says that the chance of getting heads on a toss of this coin from a certain device is a disposition of the coin and device to have such-and-such a long run frequency of heads which, in turn, is what long run frequency would be. In a key—but I think confused—passage ([1957], p. 67), Popper argues that every experimental arrangement is 'liable to produce' a sequence of frequencies, to be called probabilities, which, because they depend on the arrangement, are a property of the arrangement, namely its liability, disposition or propensity to give rise to these frequencies. Thus, an experimental arrangement of bombarding a two slit barrier with electrons is liable to give a different spread on a screen held behind the barrier from an experimental arrangement of bombarding a one slit barrier. This liability is identified with the likelihood of getting such a spread.

---

[1] Levi in his [1967], pp. 198, 203, speaks ambiguously of 'the probability disposition of a system (the coin)'. Braithwaite [1957] suggests the intermediate position that probability is a relational dispositional property, e.g. 'a property of the coin with respect to a coin throwing situation'.

[2] I understand that Mellor has modified his earlier views in his [1971]. He would say that some probabilities are not chances and need not, therefore, be related to propensities.

If, on the other hand, the propensity is—as the examples given by Popper, Mellor and Levi suggest—such a propensity as being fair, unbiased, regular or loaded, then it is related to probability as *being liable* is in fact related to *being likely*, that is, as a ground of the probability. This relation gives no clue to the analysis of probability or of being likely. What these dispositional properties explain is not the notion of probability but the notion of being equally or unequally probable. Furthermore, it is difficult to see what propensity would account for the probability of my living to be 80 or for the chance of the Conservative Party's making a 50 per cent gain in the next local elections, not to mention the probability that Tennyson wrote such-and-such a poem. The introduction of, e.g. 'A long lived virtue' or 'A local election gain virtue' would neither explain anything nor be plausible in itself. A liability to F may, though it need not, make something likely to V; but being likely to V is certainly not being liable to F.

(5) A last, perhaps less important, difficulty in the interpretation of this theory is to discover how wide its application purports to be. Its supporters from Peirce to the present day restrict it to statistical examples such as the throwing of coins or dice or the drawing of balls from an urn. Peirce, e.g. explicitly says in his [1910] that 'probability never properly refers immediately to a single event, but exclusively to the happening of a given kind of event on any occasion of a given kind'. Though Popper expressly introduces it to cover the probability of a single (or 'singular') event, e.g. 'of obtaining a 6 in the 3rd throw made after 9 o'clock this morning with this die' or of 'a single electron taking up a certain state' he identifies its spheres of application by quantum mechanics and coin tossing. Braithwaite ([1957], p. 79; cf. also Popper [1967], p. 31) explicitly rules out its application to such a thing as the probability of Hyperion winning the Derby. The theory, in Peirce, Popper and Hacking, stems from a frequency theory in which the probability of something's happening is identified with the frequency with which that type of thing happens in an actual or possible series. The new theory equates the probability not with this frequency but with an alleged propensity to produce this frequency. Popper calls it a 'revised or reformed statistical interpretation'. Even if—which I have denied—the probability of statistical examples could be analysed or interpreted in terms of a propensity, the theory has no plausibility beyond this field. It is indeed difficult to imagine what propensity

of what could explain, much less be identified with, the probability of my living to be 80, of Tennyson's having written such-and-such a poem or of Hyperion's winning the Derby. Without independent reasons for supposing 'probability' to be ambiguous, this narrowness of application is a serious defect in the theory.

In trying to discover the exact meaning of the propensity theory of probability, I have frequently alluded to the contrast between *being liable* and *being likely*. I now want to suggest that one way of characterising the fallacy of the propensity theory of probability is as an assimilation of *likely* to *liable*. Indeed, there is some evidence (cf. e.g. Popper [1957], p. 67) to suggest that such an assimilation is one root of the fallacy.

This assimilation is *prima facie* plausible for some examples, e.g. 'The coin is likely to fall heads' and 'The coin is liable to fall heads', 'It is likely to rain' and 'It is liable to rain'. But under a closer examination its plausibility vanishes.

Except in the form 'likely to V' and 'liable to V', 'likely' and 'liable' are grammatically quite unrelated. We can have a 'likely' or 'unlikely' so-and-so, e.g. explanation, consequence or candidate, but neither a 'liable' or 'unliable' so-and-so. Something can be very, most, pretty or hardly likely, but not in these ways liable. There can be a 'likelihood', but not a 'liability' of such-and-such's happening. On the other hand, someone can be 'liable', but not 'likely' to something, e.g. to prosecution, super tax, misunderstanding or fainting fits. Grammatical differences between 'liable' and 'likely' appear even in the infinitive form. 'He is likely to V' means 'It is likely that he will V', but there is no form 'It is liable that he will V' corresponding to 'He is liable to V'. The form 'likely to V' applies to any verb V, but the form 'liable to V' is severely restricted. One can say, e.g. 'He is likely to', but not 'He is liable to' with 'mean', 'intend', 'be able', 'deserve', 'know', 'solve', etc.

Logically the two notions are quite different. This is why what something is liable to V is a *sort* of thing, e.g. to rain in June, to lose five minutes per day, to fly into a rage, to fall heads, whereas what something is likely to do maybe either a sort of thing, such as any of these, or a *specific* thing, such as to rain next Tuesday, to stop within the next hour, to be held this afternoon. For any V, something can be likely to V; but the kinds of things anything is liable to V are to do (e.g. run, make a mistake) to suffer (e.g. die, fall) or have happen to it (e.g. be punished, be dropped). Usually what something is liable to V is neutral or bad. Clocks are liable to keep bad time, but not liable to keep good time; a spring is liable to break under a load, but not to support a load. But there is no limit to what something is likely to V. It can be true of something that it 'used to be' liable to V; but it makes no sense to say it 'used to be' likely to V. Conversely, it can be true of something that it is likely to have Ved, but it makes no sense to say that it is liable to have Ved. Something can be likely to be liable to V, but not liable to be likely to V.

Whatever is liable to V is not necessarily likely to V, nor is what is likely to V necessarily liable to V. The best of clerks is liable, but not likely, to make a mistake, whereas the man most likely to know the answer cannot sensibly be said to be most liable to know it. *He is not likely to be able to support two wives* cannot entail *He is not liable to be able to support two wives*, since the latter is nonsense. Being liable to V may sometimes make something likely to V and, hence, knowing or believing that something is liable to V may sometimes be a good reason for believing that it is likely to V. But being liable to V and being likely to V are not the same. Hence, one cannot analyse the notion of *being likely* in terms of the notion of *being liable*. The recent attempt to analyse probability as a propensity makes just this mistake.[1]

---

[1] I am indebted to R. Harré, D. H. Mellor and R. G. Swinburne for their comments on an earlier draft.

University of Hull

REFERENCES

BRAITHWAITE, R. B. [1957]: 'On Unknown Probabilities' in S. Körner (*ed.*): *Observation and Interpretation*, 1957, pp. 3–12.
COOPER, N. L. [1965]: 'The Concept of Probability', *British Journal for the Philosophy of Science*, 16, pp. 226–38.
HACKING, I. [1965]: *The Logic of Statistical Inference.*
LEVI, I. [1967]: *Gambling with Truth.*
MELLOR, D. H. [1969]: 'Chance', *Aristotelian Society Supplementary Volume*, 43, pp. 11–36.
MELLOR, D. H. [1971]: *The Matter of Chance.*
PEIRCE, C. S. [1910]: 'Notes on the Doctrine of Chances', *Popular Science Monthly*, 44. Reprinted in Hartshorne and Weiss (*eds.*): *Collected Papers of Charles Sanders Peirce*, 2, 1932, pp. 404–414.
POPPER, K. R. [1957]: 'The Propensity Interpretation of the Calculus of Probability and the Quantum Theory', in S. Körner (*ed.*): *Observation and Interpretation*, 1957, pp. 65–70.
POPPER, K. R. [1959]: 'The Propensity Interpretation of Probability', *British Journal for the Philosophy of Science*, 10, pp. 25–42.
POPPER, K. R. [1967]: 'Quantum Mechanics without "The Observer"', in M. Bunge (*ed.*): *Quantum Theory and Reality*, 1967, pp. 7–44.

Brit. J. Phil. Sci. 23 (1972), 45–67 *Printed in Great Britain*

Discussions

QUINE'S PHILOSOPHICAL NATURALISM

Naturalism in philosophy is by no means as simple a doctrine as it appears at first sight. Because of inherent ambiguities and because of the curious turns in its recent evolution, it is not really one coherent philosophical theory. When two naturalists say the same, they do not (necessarily) mean the same. It should not be surprising, therefore, if one naturalist feels puzzled or even indignant when reading pages written by another. It seems to me that this is what many naturalists are likely to experience when reading Quine's recent exposition of and argumentation for naturalism in epistemology in Quine [1968*a*,*b*].

Let me first dwell a bit on the ambiguities and vagaries of 'naturalism'. To begin with we may say that all naturalist methodological programmes have at least in common the claim that the method of natural science is the only legitimate or appropriate method to be used in attempting to acquire knowledge of whatever kind. We might call this claim the Principle of the Unity of Method and would not find it difficult to identify some form of this principle in the methodology of nineteenth-century positivism of Mill and Comte, in the Logical Empiricist Thesis of the Unity of Science and of Physicalism, in the Common Methodology Doctrine in the philosophy of social science, history and humanities, in marxist historicism, in the anti-positivist hypotheticism of Karl Popper ('the critical method'), in the 'science of science' programme as well as in the philosophy of Dewey and of Quine. Although all these philosophical theories do have *something* in common, it is equally obvious that they differ considerably from one another. How is this possible in spite of their agreement on the Principle of the Unity of Method? The answer to this question is easy to find, once we are aware of the ambiguities in the Principle. For (1) there are several different conceptions of the method of natural science depending on the preferred philosophy, (2) 'natural science' in the Principle may be understood either inclusive or exclusive of mathematics and mathematical theories may be regarded either as different in kind (e.g. as analytic) or not from empirical theories, (3) consequently, it is not clear whether the construction of 'conceptual frameworks' for descriptive or evaluative purposes (e.g. for 'rationality', 'democracy', 'probability', 'empirical content', 'explanatory power', etc.), meaning analysis (the Socratic method) and critical arguments without the experimental part of the method of science would count or not as the use of 'the method of natural science', (4) it is not clear whether evaluations based on the standards mentioned in (3) would count as knowledge or not, (5) consequently, it is possible for some authors to advocate the naturalist methodological programme in some disciplines but not in others (e.g. in social science but not in mathematics or in epistemology or in ethics). On account of all these ambiguities the Principle of the Unity of Method is rather a matrix which generates widely divergent methodological programmes. So although presumably all naturalists—in the wide or rather multiple sense of the Principle—agree that if a problem is solvable at all it is solvable with the help of the method of science (Mehlberg [1958]), their assent is not to exactly the same claim. For the means of solving problems and therefore solvable problems are not the same in the case of a phenomenalist (behaviourist, operationalist) and in the case of an anti-positivist theoretician. Similarly, though all naturalists agree that since our language is an instrument used for certain purposes, there is no *a priori* reason why any of its component elements should not be changed to increase effectiveness, nevertheless for Ajdukiewicz ([1935]) and Carnap ([1952]) there are statements in our language which it is logically impossible to reject (since their acceptance is necessary to use the terms involved in the appropriate sense) whereas for Tarski and Quine there are no such statements. Although, again, it follows from the Principle of the Unity of Method that the same rules of empirical method are obeyed in social science, in history and in humanities in so far as these make claim to provide knowledge and that there is no room for a peculiarly humanistic method of 'empathy', 'Verstehen' or 'intuitive understanding' (in the sense of Dilthey, Spranger and Cassirer), nevertheless to Otto Neurath those rules were the principles of physicalism, to marxists the principle of historical materialism and to Karl Popper the rules of hypotheticism combined with the principle of rationality and of social engineering. Likewise, although it follows from the Principle of the Unity of Method that both psychology and linguistics use (and should use) the empirical method of natural science, to Quine this means behaviourism and to Chomsky anti-behaviourist 'rationalism'. We may conclude, therefore, that the reference in the Principle to 'the method of natural science' is deceptive for it wrongly suggests that there is a common positive standard of the method of natural science shared by all naturalists and somehow accessible to inspection in case of doubt or of dispute.

Both past and present schools of naturalism originate from the criticism of various apriorist, speculative and irrationalist trends in philosophy which underrate empirical science and claim that there is an intuitive or specifically philosophical method of acquiring knowledge, prior and superior to the method of science. Consequently all naturalists agree that there is no room for extra- or supra-scientific methods. However, although they do seem to have common foes, namely apriorists, mystics and avowed irrationalists, they rarely—if ever—close ranks to attack the same foe. Rather they are concerned with what they regard as *misconceptions of the method of science* in the philosophy of other fellow naturalists. So, for example, those who favour the sociological, psychological or historicist brands of naturalism are concerned with criticising the 'lack of realism' or 'fictitious constructs' of those who use the method of logical or rational reconstruction; they do it apparently in the belief that they fight against 'disguised apriorism'. On the other hand, those who practise 'the logic of science' feel that their opponents in the naturalist camp open the door to irrationalism by overemphasising the non-rational elements in science (conceived as behaviour), by looking exclusively for external (sociological, political, economic) determinants of scientific activity or simply by disregarding the descriptive and explanatory function of scientific theories. Accordingly, we have controversies between Kuhn and Popper and between Quine and Carnap.

This is not to say that Quine's epistemology and philosophy of science is in all respects closer to Kuhn's than to Carnap's or Popper's. Anyone who believes—in spite of Feyerabend's philippic—that empiricism (blended with conventionalist elements) is a tenable and fruitful position, will find a lot of encouragement in Quine's essays. For what could be more reassuring than his condemnation of the 'epistemological nihilism' and 'cultural relativism' of Polanyi, Kuhn and Hanson? Or more refreshingly orthodox than his claim that "... Two cardinal tenets of empiricism remained unassailed ... and so remain to this day. One is that whatever evidence there is for science is sensory evidence. The other ... is that all inculcations of meanings of words must rest ultimately on sensory evidence ..." (Quine [1968*b*], p. 73). On the other hand, Quine's identification of the naturalist programme with the reduction of epistemology (and of the philosophy of science) to behaviourist psychology and linguistics will be unacceptable to many naturalists, the more so that his argumentation for that reduction is far from convincing.

In the rest of this article I shall attempt two things. In §1 I shall examine critically Quine's argument in support of his naturalist programme (i.e. to practise epistemology as a chapter of psychology and linguistics) from the failure of the Cartesian and reductionist programmes. In §2 I shall examine the relation between Quine's statement of the thesis of naturalism according to which knowledge, mind and meaning are part of nature, and his naturalist programme in epistemology.

§1. *Quine's argument from the failure of the Cartesian and reductionist programmes*

Quine's preference for the naturalist way of practising epistemology is rationalised in Quine [1968*b*] in the form of an argument whose main premiss is the statement of the failure of the Cartesian quest for certainty and of the reductionist programme of the *Logischer Aufbau* type. In outline the argument is as follows: If one looks for certainty in mathematics and in science and if one believes that it is possible to reduce mathematics to logic and science to a phenomenalistic basis by defining less intuitive concepts with the help of intuitive ones and by deducing less obvious postulates from obvious ones (e.g. by deducing scientific theories from observations), then the method of rational (or logical) reconstruction is preferable to the methods of psychology in epistemology, since the latter is unable to perform the desired reduction and since it would be circular to appeal to science in an attempt at justifying science. However, this programme of reducing mathematics and science to indubitable elements has proved impossible. Why then '... all this creative reconstruction?' 'Why not just see how this construction (of science) really proceeds? Why not settle for psychology?' (p. 75).

Even though there is no fear of circularity '... once we have stopped dreaming of deducing science from observation' (p. 76), rational reconstruction of science would have an advantage over psychology, if it were able to establish 'the essential innocence' (p. 76) of theoretical concepts and statements by defining each of the theoretical concepts in observables and by translating each of the theoretical statements into observational statements, i.e. if rational reconstruction of the *Logischer Aufbau* type were possible. However, this programme too has proved impracticable. It has proved impossible because of the indeterminacy of translation, i.e. 'because a statement about the world does not always or usually have a separable fund of empirical consequences that it can call its own' (p. 82). 'To relax the demand for definition, and settle for a kind of reduction that does not eliminate, is to renounce the last remaining advantage that we supposed rational reconstruction to have over psychology; namely, the advantage of translational reduction. If all we hope for is a reconstruction that links science to experience in explicit ways short of translation, then it would seem more sensible to settle for psychology. Better to discover how science is in fact developed and learned than to fabricate a fictitious structure to a similar effect' (p. 78).

Unless I have misunderstood Quine completely, the last part of his argumentation, beginning with the question Why then '. . . all this creative reconstruction?' and ending with the conclusion advising us to settle for psychology, seems odd and unconvincing. For the conclusion advises us to choose one course in preference to another one, while the premisses fail to mention other alternatives; the implicit assumption that there are just two alternatives available is glaringly false; finally, it is implied that we have a choice to make which in fact we do not have.

Let me restate briefly and in a schoolboy fashion the relevant part of the argumentation. It seems to be as follows: *(1)* Shall we choose rational reconstruction of the *Logischer Aufbau* type or psychology? *(2)* If translational reduction of the *Logischer Aufbau* type is possible, than it is preferable to psychology. *(3)* If all we hope for is to link science to experience in a way short of translation, then psychology is preferable. *(4)* But translational reconstruction of the *Logischer Aufbau* type is not possible because of the indeterminacy of translation. Hence *(5)* Psychology is preferable.

Question *(1)* is, of course, where all the trouble begins. For it falsely assumes that only one of only two direct answers is to be chosen, whereas one may want to reject both and examine other alternatives or to accept both, together with some other alternatives perhaps. So question *(1)* is at the same time like 'Have you stopped beating your wife?', since it excludes or ignores other alternatives, and also like 'Shall I clean my teeth or else wash my hands?', since it creates an unnecessary dilemma. But if we refuse to include among our premisses the false assumption on which question *(1)* rests, then the argument becomes an obvious *non sequitur*.

Surely from the failure of the Cartesian quest for certainty and of the reductionist programme of the translational type it does not follow in any clear sense that 'it would seem more sensible to settle for psychology'. *If* we want to establish 'the essential innocence' of theoretical terms and know that piecemeal translational reduction is impracticable on account of the indeterminacy of translation, then we may settle for some elimination programme of either Ramseyan or Craigian type.

The Ramsey sentence of a theory T is a theory T^R obtained from T by second order existential generalisation on all theoretical terms of T. So T^R no longer contains any theoretical terms of T and yet is observationally equivalent to T, i.e. for every observational sentence O: $T \vdash O$ if and only if $T^R \vdash O$. (Ramsey [1931], Carnap [1966], Rozeboom [1963], Tuomela [1969].) It is obvious, therefore, that the elimination of theoretical terms with the help of Ramsey sentences does not amount to translational reduction of the definitional type. If a theoretical term is explicitly definable, then it is Ramsey-eliminable, but the converse does not hold. Similarly, Craig's elimination programme, i.e. the replacement of theories with auxiliary terms by observationally equivalent theories without auxiliary terms (Craig [1953], [1956], [1960]; Hintikka [1964]) does not rely on translational reduction and so is not affected by the indeterminacy of translation, i.e. by Quine's principal criticism of the reduction of the *Logischer Aufbau* type. It should be noted that many criticisms of Craig's elimination method (Black [1967], Hempel [1958], Nagel [1961]) are irrelevant from the point of view of the epistemological question under discussion here. For they are concerned with adverse effects the elimination of theoretical terms would have on actual scientific theories in terms of systematic economy, ability to grow, etc. But in establishing the innocence of theoretical terms by appeal to their eliminability in principle with the help of Ramsey's or Craig's method, we are no more committed to the elimination of theoretical terms from actual scientific theories than we are to the elimination of all defined terms from actual mathematical theories by showing that they are all eliminable in principle. Anyway, since the elimination of theoretical terms is possible without translational reduction (and is even mentioned by Quine in his [1968*b*], pp. 79–80), it is not true that if piecemeal translational reduction of the early Carnapian type fails, then the only type of reduction that remains is the one which does not eliminate theoretical terms and which—according to Quine—does not have any superiority over psychology.

But where has Quine shown that translational reduction is the only advantage of rational reconstruction over psychology? Where are his arguments in support of this claim? Surely it does not follow from the fact that psychology cannot provide translational reduction. In other words: 'If we are hoping for translational reduction, then only logical analysis is worth pursuing (since psychology is unable to provide it)' does neither imply nor support in any clear sense the statement 'Logical analysis is only worth pursuing if we are hoping for translational reduction'. Not only are there no arguments in Quine's article in support of this claim but an obvious sense of 'establishing the innocence of theoretical terms' is ignored which—if taken into account—would have shown the claim to be false. The primary concern with theoretical terms, felt by empiricists, is their liability to be used in formulating untestable knowledge claims which, when mistaken for determinate factual statements, become the subject of futile controversy. To show a set of concepts innocent in this respect it would indeed be *sufficient* to show that every statement in which any occurs is completely translatable into one in which none occurs, i.e. into observational statements. But it is by no means necessary. For it would also be sufficient (*a*) to make explicit the rules by which theoretical terms acquire denotations and (*b*) to adopt the convention that their denotations and the truth values of statements in which they occur are no more determinate than the rules ensure, as is done in the model theoretical treatment of empirical theories with theoretical terms (Przelecki [1969]). This account of empirical theories is not only compatible with Quine's thesis of the indeterminacy of translation but gives an exactly corresponding result. More particularly, when the problem of 'translation' is between a theoretical and observational language, namely to find logical correlations (not necessarily of the definitional sort) between statements of the two languages, a variety of solutions is now known to be possible. This indeterminacy is even strengthened by the fact that the model-theoretical reconstruction of empirical theories used so far presupposes the analytic-synthetic distinction. Far from making the logical analysis redundant or uninteresting, these results make it all the more necessary to undertake a logical inquiry into the consequences of adopting one of the many solutions (Przelecki [1969], Przelecki and Wojcicki [1969], Williams [1970]). Needless to say, it would be circular to look to psychology for criteria of empirical meaningfulness of theoretical terms in all scientific theories (including psychological theories).

We are told by Quine that it is "better to discover how science is in fact developed and learned than to fabricate a fictitious structure to a similar effect". Does the same apply to mathematics? Strictly speaking, since the Cartesian and the reductionist programme have failed in mathematics also, should Quine not have argued that we should give up logical research into the foundations of mathematics and devote our attention exclusively to the psychology of mathematics? Admittedly, there has been more progress in the logical studies of mathematics than in the logical studies of science, but if this is taken as an argument against the feasibility of a logic of empirical theories, then an exactly analogous argument should convince us that it is not reasonable to hope for an empirical science of human behaviour; and then it would seem more sensible to settle not for psychology but for physics, physiology and biochemistry.

Let us look at some of Quine's own essays on epistemological matters, for example, at Quine [1968*a*]. The principle of ontological relativity according to which 'it makes no sense to say what the objects of a theory are, beyond saying how to interpret or reinterpret that theory in another' (p. 50), seems to owe more to Quine's interest in the logical foundations of mathematics than to his interest in experimental psychology or in linguistics. And of course all the relevant results of proof and model theory to which he refers in his most interesting discussion of ontological relativity are based on 'creative reconstruction' and on 'fictitious structures' of formalised mathematics rather than on a study of how mathematics "is in fact developed and learned". Why should not epistemology then be a chapter of meta-mathematics or of metascience in general?

Of course, rational reconstruction need not be associated with a formal background of a system of logic and meta-logic. It may be of an 'understanding' type or else of decision-theoretical nature. In the former case one reconstructs the activities of scientists (as in Karl Popper's *Logic of Scientific Discovery*) or of mathematicians (as in Imre Lakatos's *Proofs and Refutations*) as rational, i.e. purposeful behaviour, as if these activities were designed to achieve certain aims, postulated as the aims of science or of mathematics. In the latter case, our paradigm is statistics or decision theory or game theory and we are engaged in constructing postulates for 'rational decision making'; Our theories are then normative or else 'ideal types'. One may suggest that there is no essential difference between this type of rational reconstruction and theoretical psychology or linguistics. For what we are offered by theoretical psychologists and linguists are just highly simplified and idealised models, usually with the proviso that these models are to be regarded as instruments rather than as true or false descriptive statements. Well, if there is no difference at all, then what is there to naturalise in epistemology? If, on the other hand, there is a difference, then we have here an alternative to psychology (or rather a complementary discipline).

It seems, however, that had all these alternatives to psychology been taken into account, Quine's argument would not have improved appreciably. For it would still be concerned with the question whether to choose psychology of science or any other alternative. But why not do *both* the logic and psychology of science? Why not study both the logical foundations and the psychological processes of mathematics? Why not encourage both formal and empirical semantics? Unless we are concerned with the practical problem of how best to allocate scarce resources among competing research programmes in our universities or with the individual choice problem how best to spend one's limited time, there does not seem to be any good reason why a choice between psychology and logical reconstruction (applied logic) is to be made at all. It seems to me an unfair account of the situation in the philosophy of science and in epistemology to claim either that a struggle for survival has to go on between two rival 'paradigms'—the traditional and the naturalist—one of which is bound to replace the other, or to claim that a dramatic choice is to be made between two programmes only one of which is scientifically respectable. It would seem more plausible to say with Carnap that:

> ". . . The task of analyzing science may be approached from various angles . . . (Some of) these investigations of scientific activity may be called history, psychology, sociology and methodology of science. The subject matter of such studies is science as a body of actions carried out by certain persons under certain circumstances . . . it is certainly an essential part of the foundations of science. We come to a theory of science in another sense if we study not the actions of scientists but their results, namely, science as a body of ordered knowledge. Here by 'results' we do not mean beliefs, images, etc., and the behaviour influenced by them. That would lead us again to psychology of science. We mean by 'results' certain linguistic expressions, viz. the statements asserted by scientists. The task of the theory of science in this sense will be to analyze such statements, study their kinds and relations, and analyze terms as components of these statements and theories as ordered systems of those statements . . . It is possible to abstract in an analysis of the statements of science from persons asserting the statements and from the psychological and sociological conditions of such assertions. The analysis of the linguistic expressions of science under such an abstraction is logic of science" (Carnap [1938], pp. 408–9).

§2. *Quine's naturalist thesis and his naturalist programme in epistemology.*

Quine formulated briefly the main tenets of naturalism in the following passage:

> "... Philosophically I am bound to Dewey by the naturalism that dominates his last three decades. With Dewey I hold that knowledge, mind and meaning are part of the same world that they have do do with, and that they are to be studied in the same empirical spirit that animates natural science. There is no place for a prior philosophy" (Quine [1968*a*], p. 26).

The quoted sentences seem to express a *prima facie* descriptive thesis ("... knowledge, mind and meaning are part of ..."), a normative thesis or programme ("... they are to be studied ...") and another normative thesis ("There is no place for a prior philosophy"). If the first thesis denies that knowledge, mind and meaning belong to a world entirely different from the world with which scientific (and mathematical) knowledge are concerned, the third thesis denies the existence of knowledge which is prior and superior to scientific knowledge and the naturalist programme recommends that we study knowledge, mind and meaning with the help of scientific method in a very broad sense which includes the methods of science, of mathematics, of logic, decision theory, social science and history, etc., then I do not want a dispute with naturalism so understood. For the first thesis seems too obscure and I sympathise with the—admittedly also vague—claims of empiricism and the unity of knowledge and method expressed by the remaining two claims. If, however, the first thesis is claimed to give support to the programme of reducing epistemology to behaviourist psychology and linguistics, then it may be worth a dispute. Furthermore, we may want to qualify the claim that there is no room for a prior philosophy by reference to the well-known criticism of comprehensive rationalism and by reference to the conventional elements in science.

The following seems a fair though pedantic reconstruction of an argument which purports to use Quine's naturalist thesis in justifying the naturalization of epistemology in Quine's sense:

> (*1*) Knowledge, mind and meaning are properties and dispositions of man and of his behaviour, (*2*) Man and his behaviour are natural phenomena, (*3*) Since knowledge, mind and meaning are the subject of epistemology, therefore the subject of epistemology is a natural phenomenon, (*4*) The only method appropriate in the study of natural phenomena is the empirical method of science. Therefore, (*5*) The only method appropriate in the study of knowledge, mind and meaning is the method of natural science applied to human behaviour (i.e., the method of experimental psychology, behavioural linguistics, etc.).

It seems intuitively obvious that this argument is not valid unless premisses (*1*) and (*2*) are strengthened with a proviso such as "... are nothing but ..." and unless an additional premiss is introduced asserting that the methods of experimental psychology and behavioural linguistics are identical with the methods of science applied to human behaviour. Without the mentioned proviso, premisses (*1*) and (*2*) are acceptable to many anti-naturalists since—as these premisses stand—they do not deny that knowledge, mind and meaning may be something more than properties and dispositions of man and of his behaviour. Accordingly, it does not follow from these premisses in the weaker sense that the only appropriate method in the study of knowledge, mind and meaning is the method of natural science (whatever it is).

But if the stronger interpretation is placed on premisses (*1*) and (*2*), would these premisses be true? How would we argue about their truth with anti-naturalists or with anyone who denies them? Are these premisses meaningful statements? Are they not making absolute ontological claims which on Quine's principle of ontological relativity are without sense?

Suppose that we are confronted by the following series of apparent counterclaims:

'Knowledge, in the form of scientific and mathematical theories belongs to the world of ideas and is to be studied not with the methods of psychology but by analyzing, developing and criticising the content of theories in relation to problems which are likewise susceptible to similar scrutiny . . .' Is the above view incompatible with Quine's naturalism? What kind of controversy is this (if it is a controversy)?

It seems that the sentences 'Knowledge, mind and meaning are properties and dispositions of man and of his behaviour' and 'Knowledge in the form of scientific and mathematical theories belongs to the world of ideas' are best interpreted not as incompatible, descriptive statements or hypotheses but as expressing epistemological decisions or as expressing proposals for different research policies. Sentences (*1*) and (*2*) in the naturalist argument clarify what it means to be a behaviourist or what it means to speak his language when discussing 'something like' traditional epistemological problems; they are perhaps meaning postulates of the behaviourist language and, if so, have to be accepted as true if one wants to use the behaviourist's language in epistemology; naturally, to decide to use the behaviourist language in epistemology amounts to a decision to study epistemological problems in behaviourist terms. So what looked like an argument in support of Quine's naturalist programme in epistemology turns out to be nothing but clarification of the position of naturalism in its behaviourist form and decision to study the naturalist counterparts of traditional epistemological problems. Sentences (*1*) and (*2*) in the argument are unable to support conclusion (*5*) (i.e., the naturalisation of epistemology in Quine's sense) against either anti-naturalists or other critics who do not share the behaviourist assumptions or preferences. If by classifying knowledge as either part of the world of nature or as part of the world of ideas we merely propose to study different problems with the help of suitable methods, then the results of such two research policies need not be incompatible but may be complementary. If, on the other hand, by choosing one we also condemn the other as not worthy of the status of 'knowledge' and of 'scientific method', then we commit ourselves thereby to a corresponding evaluation of the outcomes of the two approaches to epistemology. Now, if decisions like (*1*) and (*2*) or their alternatives are unavoidable in epistemology, in the philosophy of science and in science itself, then does this not suggest that there has to be a place for 'something like' a prior philosophy?

Finally, is Quine's programme itself unambiguously given by reference to 'the empirical spirit' or to 'the empirical method of science'? There are various apparently incompatible accounts of the empirical method of science given by different schools in epistemology and in the philosophy of science. Which one do naturalists choose? and how do they choose it? Do we use the empirical method of natural science to make the choice? But by which account of the empirical method do we choose which account of the empirical method to accept? This approach seems to involve us in a circle, so we must try a different one. Without relying on any account of scientific method we may simply try to learn it from scientists and then use it in accordance with the naturalist programme. But who do we learn it from? From Mach or from Einstein? From zoologists or from astronomers and cosmologists? Why not from astrologers and scientologists? Are pure mathematicians scientists in this terminology? What is wrong with the empirical spirit that animated Carnap and other Logical Empiricists? What is wrong with non-behaviourist schools in psychology and in linguistics? Are not answers to these questions given by naturalists and by their opponents 'prior philosophies' which cannot be avoided however scientifically minded one may be? (Cf. Popper's criticism of comprehensive rationalism in Popper [1945]; also Lakatos [1972].)

Although reference to any account of scientific method in deciding between naturalism and its alternatives has just been dismissed as involving us in a circle, it seems appropriate—in concluding these critical comments—to mention Quine's own account of scientific method. Elaborating Duhem's point that single hypotheses are never falsified because consequences are usually drawn from clusters of statements, Quine emphasised that it is always possible to save one's hypothesis by making suitable adjustments elsewhere in one's science (including mathematics and logic). Whether to do so or not amounts to a decision which has not been dictated by experimental findings themselves (the decision not to use conventionalist strategems, as Popper called it). Now surely this applies *a fortiori* to naturalised philosophy; it would be absurd to expect philosophical doctrines to be more sensitive to testing than scientific theories. If so, we can never be forced to reject or accept any doctrine in epistemology or in the philosophy of science by scientifically studying the scientist's behaviour, present or past; ultimately the choice depends on our decision (Cf. Lakatos [1972]). At least these decisions, on Quine's own account of scientific method, may be regarded as irreducible *a priori* components of both science and of philosophy.

JERZY GIEDYMIN
The University of Sussex

REFERENCES

Ajdukiewicz, K. [1935]: 'Die Wissenschaftliche Weltperspektive', *Erkenntnis*, **5**, pp. 22–30. Reprinted in English in H. Feigl and W. Sellars (*eds.*): *Readings in Philosophical Analysis*, 1949, pp. 182–188.

Bergmann, G. [1943]: 'Outline of an empiricist philosophy of physics', *American Journal of Physics*, **11**, pp. 248–58, 335–44. Reprinted in H. Feigl and M. Brodbeck (*eds.*): *Readings in the Philosophy of Science*, 1953, pp. 262–88.

Black, M. [1967]: 'Craig's Theorem', in P. Edwards (*ed.*): *The Encyclopedia of Philosophy*, **2**, 1967, pp. 249–51.

Carnap, R. [1938]: 'Logical Foundations of the Unity of Science', in the *International Encyclopedia of Unified Science*, **1**, pp. 42–62. Reprinted in H. Feigl and W. Sellars (*eds.*): *Readings in Philosophical Analysis*, 1949, pp. 408–423.

Carnap, R. [1952]: 'Meaning Postulates', *Philosophical Studies*, **3**, pp. 65–73.

Carnap, R. [1966]: *Philosophical Foundations of Physics*.

Craig, W. [1953]: 'On axiomatizability within a system', *The Journal of Symbolic Logic*, **18**, pp. 30–32.

Craig, W. [1956]: 'Replacement of auxiliary expressions', *Philosophical Review*, **65**, pp. 38–55.

Craig, W. [1960]: 'Bases for first order theories and sub-theories', *Journal of Symbolic Logic*, **25**, pp. 97–142.

Feigl, H. [1949]: 'The Scientific Outlook: Naturalism and Humanism', *American Quarterly*, **1**. Reprinted in H. Feigl and M. Brodbeck (*eds.*): *Readings in the Philosophy of Science*, 1953, pp. 8–18.

Hempel, C. [1958]: 'The theoretician's dilemma', in H. Feigl, M. Scriven and G. Maxwell (*eds.*): *Minnesota Studies in the philosophy of science*, **2**, pp. 37–98. Revised and reprinted in C. Hempel: *Aspects of Scientific Explanation*, 1965, pp. 173–243.

Hintikka, J. [1964]: 'Distributive normal forms and deductive interpolation', *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **10**, pp. 185–91.

Lakatos, I. [1972]: 'History of science and its rational reconstructions', in R. C. Buck and R. S. Cohen (*eds.*): *Boston Studies in the Philosophy of Science*, **8**, pp. 91–136.

Mehlberg, H. [1958]: *The Reach of Science.*

Nagel, E. [1961]: *The Structure of Science.*

Popper, K. [1945]: *The Open Society and Its Enemies.*

Przelecki, M. [1969]: *The Logic of Empirical Theories.*

Przelecki, M. and Wojcicki, R. [1969]: 'The Problem of Analyticity', *Synthese*, **19**, pp. 374–99.

Quine, W. [1968*a*]: 'Ontological Relativity', *Journal of Philosophy*, **65**, pp. 185–212. Revised and reprinted in W. Quine: *Ontological Relativity and Other Essays*, 1969, pp. 26–68.

Quine, W. [1968*b*]: 'Epistemology Naturalised', (invited address at the 19th International Congress of Philosophy). Reprinted in W. Quine: *Ontological Relativity and other Essays*, pp. 69–90.

Ramsey, F. [1931]: *The Foundations of Mathematics and Other Logical Essays.*

Rozeboom, W. [1963]: 'The factual content of theoretical concepts', in H. Feigl and G. Maxwell (*eds.*): *Minnesota Studies in the Philosophy of Science*, **3**, pp. 273–357.

Tuomela, R. [1969]: *Auxiliary concepts within first-order scientific theories.* (Ph.D. Dissertation, Stanford University.)

Williams, P. [1970]: 'On the Logical Relations between Expressions of Different Languages' (paper read at the 1970 Annual Conference of the British Society for the Philosophy of Science in Cambridge).

EPICYCLIC POPPERISM

1. The great merit of Popperism is that it challenged hard-line Empiricism, as represented by the positivists, well before the general decay of that position set in. The fallacious and always unsatisfactory doctrine of enumerative (or 'ampliative') induction as typical of empirical scientific method was firmly countered and set aside. This was clear gain, and besides, Popper made numerous, if often oblique, epistemological suggestions which pointed to a much more fruitful approach in the philosophy of science (see my [1970], Ch. IV, iii.). But Sir Karl was still held by the spell of Empiricism sufficiently to seek a criterion of demarcation between science and metaphysics in empirical falsifiability, which entangled him in some of the very difficulties which his reforms sought to remedy.

The king-pin of the theory was the unassailable logical validity of *modus tollens* and the possibility of complete valid refutation of a universal proposition by a contradictory particular. Scientific theories, therefore, though incapable of inductive verification short of progression *ad infinitum*, seemed conclusively refutable by a single instance. But the single instance had to be intersubjectively observable and the falsifier in the last resort must be a 'basic' statement.

Admission by Popper that no single observation was really acceptable as a refutation, unless repeated and checked, involved him in a new, though surreptitious and veiled, appeal to some sort of inductive confirmatory procedure, for how repetition could function logically as validating was not apparent. And though he clearly saw that observation and interpretation were always mutually involved (cf. Popper [1959], p. 94 f.), he persisted in requiring basic statements as ultimate falsifiers, and thus failed in the final issue to free himself from the weaknesses of Empiricism. (Cf. my [1970], pp. 75 ff.) Indeed, he could hardly do otherwise if he was to preserve intact the logical foundation of his whole theory—the impregnable validity of refutation by a singular negative instance.

A second weakness in the theory was that while it represented the progress of science as a series of 'conjectures and refutations', it left the source of the conjectures unexplained and held them to be logically inexplicable (relegating any investigation of their origins to the psychologist). The scientist thus would seem to have no guide in his selection of new conjectures apart from his own scientific intuition the nature of which remains a mystery for the philosophy of science. (Cf. my [1970], p. 80.)

The theory, however, became dubious as soon as N. R. Hanson announced his rediscovery of what had long since been maintained in other contexts, by philosophers like Bernard Bosanquet, Harold Joachim and Brand Blanshard, as well as by *Gestalt* and other psychologists. Hanson drew attention to the fact that perception is never of bare or raw data but is always of objects interpreted in the light of funded background knowledge, so that all observation is 'theory-laden'.[1] And the whole position was called in question when Thomas Kuhn produced his theory of scientific revolutions as exceptional crises in the history of science, occurring between long periods of 'normal' science. Science, Kuhn maintained, could proceed only under the aegis of a 'paradigm' or conceptual scheme which was not only in large measure immune from refutation by particular anomalous instances, but even determined the way in which observations were made and how its results were understood (or indeed were actually 'seen'). (Kuhn [1962].)

In the light of these new, influential and upsetting theories some of Popper's follower's have tried to refute Kuhn more or less *a priori* by denying the existence of 'normal' science. (Cf. Watkins [1970].) But others, eminently Imre Lakatos, have attempted to adapt falsificationism to conform to the historical facts adduced by Kuhn as evidence for his position. What Lakatos does, in effect, is to add a series of epicycles to the earlier form of Popperism, to produce a theory that becomes progressively more complicated without finally overcoming the internal inconsistencies.

2 The first epicycle Lakatos descries lying implicit in the original Popperism.

Originally Popper rejected conventionalism, not because it was logically unsound, but because it was methodologically unprogressive and thus 'unscientific'. He recommended the methodological rule that no theory should be held by convention to be immune from refutation by counter-instances. But then, finding no genuine basic statements available to serve as falsifiers he was constrained to assert that we adopted certain statements (or classes of statements) as basic by convention in order to keep our speculations scientifically respectable. (Popper [1959], § 29.) Accordingly, Lakatos identifies the original form of the doctrine as 'dogmatic' (or 'naturalistic') falsificationism, and adds (or attributes to Popper himself the addition of) the first epicycle. A form of conventionalism is now admitted as scientifically respectable and 'methodological falsificationism' is introduced. (See Lakatos [1970], pp. 103 ff.) This makes basic statements 'basic' by convention, depending on the recognition of an accepted experimental technique, which is accepted also by convention. For the scientist realises that any such technique involves theoretical assumptions, many if not all of which are fallible. But these are relegated to what is called the 'unproblematic background of knowledge'. Thus a statement is 'observational' by convention also, and Empiricism recedes a step further.

---

[1] See N. R. Hanson [1958], chs. i and ii; B. Bosanquet [1911], I, ch. i, p. ii; H. H. Joachim [1948], ch. ii; B. Blanshard [1939], I, bk. i and *passim*. Popper himself had, of course, drawn attention to this characteristic of observation in his [1935] but never adequately developed the point or its implications for his own doctrine.

The new position, however, is anything but free from difficulties, some of which Lakatos recognises. It requires yet more conventional concessions and rules of procedure. There must be rules (it seems arbitrary) for the falsification of statistical theories which are proof against refutation by single instances. There must also be decisions as to what saving clauses invoking 'ceteris paribus' may, and what may not, be included in the unproblematic background of knowledge. Again, some criterion is needed (and lacking) for demarcating the unproblematic background from the theory to be tested; and this is a more crucial matter than Lakatos (or Popper) seems to allow. For, a potential falsifier, to be acceptable as such, must not only be checked by repetition (preserving the entanglement with induction), it must also be supported by 'a well corroborated falsifying hypothesis,' (cf. Lakatos [1970], p. 108, and Popper [1959], § 22) the precise status of which remains obscure.

For instance, Lavoisier's experiments with mercury might have been described by Priestley as follows: On heating, the mercury absorbs a portion of the air and gives off phlogiston, which is less in volume than what is absorbed, but is sufficient to render the remaining air incapable of supporting combustion. When the calx is heated, phlogiston is reabsorbed and so the remaining air becomes dephlogistigated and 'highly respirable'. The phlogiston hypothesis might be (was by Priestley) regarded as well corroborated, and thus the observation statements could be held to refute Lavoisier's hypothesis. On the other hand, the refutation of the phlogiston hypothesis only occurs if the observations are interpreted as supported by Lavoisier's view that some part of the air (the respirable part) is fixed during calcination and that this is given off again during reduction. Thus the supporting hypothesis would seem to be the one actually being tested—a complete subversion of the Popperistic view.

Further, methodological falsification gives no more clue to the source and origin of new hypotheses than does dogmatic falsificationism. But this, we shall find, is a defect[1] persistent in all forms of Popperism.

In any case, as Lakatos points out, the whole position collapses in the light of historical evidence. The Newtonian theory tolerated far too long such anomalies as the perihelion of Mercury. Galileo accepted the Copernican theory in the face of what at the time seemed incontrovertable contrary evidence. Bohr accepted a conception of light emission which ran counter to Maxwell's well corroborated theory. In actual historical fact, the contest is not just between proposed theory and experimental result, but between two theories (at least), of which the experimental result refutes one while, and because, it confirms another.

---

[1] Considered, apparently, to be a merit by some Popperians.

3 Lakatos, accordingly, introduces a second epicycle, and the new version is called 'sophisticated methodological falsificationism'. According to this version, falsification can occur only through the corroboration of a new and better theory. Two new rules are involved, one for accepting any theory as scientific, and one for rejecting a theory as falsified. They are really the same rule seen, as it were, from opposite ends.

First, no theory is to be acceptable which does not (*a*) explain all the facts explained by previous theories and (*b*) predict new facts to boot, some at least of which have been corroborated. Secondly, a theory is to be rejected if, but only if, another rival theory has proved itself capable of doing both (*a*) and (*b*) above. Moreover, theories are never appraised in isolation, but always together with their initial conditions and auxiliary hypotheses (one might almost consider these provisions as further epicyclic complications[1]); and they can only be judged in series such that the later ones advance in the manner described beyond the earlier ones. Each successive theory adds auxiliary clauses to, or involves semantic reinterpretations of, the previous theory in order to accommodate some anomaly; each has at least as much unrefuted empirical content as its predecessor; and each predicts novel facts which its predecessor failed to anticipate (or even forbade). Such a series is called a problem-shift, and it is progressive so long as the new facts predicted in each case are verified or corroborated,[2] at least to some extent. Falsification now has a 'historical character'. It is no longer conceived as a simple relation between a theory and its empirical basis, but as 'a multiple relation between competing theories, the original "empirical basis" and the empirical resulting growth from competition'.

The new epicycle has the alleged effect of reducing the conventional element in the theory, but what remains unavoidably conventional is what is to count as 'fact', what experimental technique is to be accepted as revealing the fact, and what interpretative hypotheses are to be adopted as determining what the fact is; as well as what is to count as refutation of statistical or probabilistic theories.

It may be noted immediately that we have moved a long way from 'dogmatic falsificationism' and it is now hard to see what has become of the central validating principle of the method, that universal statements can be conclusively falsified by singular ones, though never conclusively verified. It has, in fact, been abandoned, for Lakatos says that falsification is not a diadic relation between theory and fact, and that 'no experiment, experimental report, observation statement or well-corroborated low-level falsifying hypothesis alone can lead to falsification'. ([1970], p. 119.)

Moreover the 'empirical basis', though it is still identified with the 'scientific character' of theories and is now said to be inseparably connected with theoretical progress, has become marvellously transformed. Taken at its face value, the account so far given of scientific procedures would seem to be a return to the notion, so emphatically rejected by Popper, of scientific advance by accumulation of corroborated (verified, or proven) fact. But to judge it so would be to overlook the consequences of its conventional elements. For what counts as a 'fact' 'depends on our methodological decision' (op. cit. p. 129) and this involves the adoption of 'interpretative theories'. So the decision how to proceed when theories apparently clash with facts, whether to ignore anomalies or to abandon the theories they belie, is really a question whether to maintain 'explanatory theory' in preference to 'interpretative theory' or vice versa. Or better, it poses a problem of which theory to regard as interpretative and which as explanatory. Clash between 'theory' and 'fact', therefore, is really nothing more nor less than *inconsistency* between theories. (Lakatos [1970], p. 130.) And a progressive problem-shift is a progressive adjustment of theories so as to remove inconsistencies, and one which, at the same time, increases the corroborated content.

---

[1] As Lakatos develops these provisions into his theory of the methodology of scientific research programmes, with their 'negative' and 'positive' heuristics, the analogy becomes still more appropriate.

[2] Lakatos ([1970], pp. 120–1) uses the terms synonymously.

4 Let us pause here to take stock of the new position. Popper's original aim was to provide a criterion of scientific method that would distinguish it clearly from metaphysics, and yet avoid the fallacies and futility of crude positivism. To do so he felt it necessary to preserve some empirical basis for theory, and as induction was unacceptable, for the various reasons he set out, he opted for falsifiability by basic empirical statements. The difficulties of this position have led Lakatos to modify it epicyclically until the empirical basis has become transformed by a series of conventional decisions, and falsification now consists in the disclosure of inconsistency between theories (or hypotheses).

In the light of this transformed notion of 'empirical basis', what can now be understood by 'corroboration'? To be acceptable a theory must explain all (or most of) the old facts and predict new ones, some at least of which must be corroborated. How is this last to be done? For Popper corroboration of a theory simply meant the continued failure of falsification, but there was always difficulty in applying the word to *facts*. The only appeal which seemed possible here was to direct observation. But while Popper hedged direct observation about with provisions like repeatability, inter-subjectivity, and other conditions which we need not further specify, Lakatos now tells us plainly that isolated observation statements cannot falsify and that what is a fact is a matter of convention depending upon interpretative theory. We have already noted the difficulty of deciding whether that interpretative theory is to be the prior hypothesis which is being superseded, the new one which is its rival, or some other the relation of which to either remains obscure. How then can 'facts' be corroborated, unless, once more, appeal is surreptitiously being made to some discredited form of inductive procedure?

We may press the matter further and ask how facts, new or old, are predicted by a hypothesis. Presumably an observation statement must be deducible from it (along with initial conditions) which experiment will bear out. But the existence of the initial conditions, the experimental technique, the theories implicit in it and the interpretation of the results, are all matters of convention. No clear explanation is forthcoming of the relations between explanatory theory and interpretative theory, nor how these again are related to experimental techniques and the conventional decisions required for scientific research. Accordingly, if all the important decisions are made in conformity to the earlier theories that the new hypothesis is proposed to replace, its defeat will be a foregone conclusion. If, on the other hand, the conventions are favourable to the new hypothesis it must always be corroborated. What now has become of our demarcation criterion distinguishing science from metaphysics?

The question persists: What is the source of new remain mysterious inspirations of the researcher. impenetrable obscurities. If isolated observation theory, what is the source of anomalies? And what seeking a new theory? Naive falsification presume proved by a conflicting observation, a new hypothe now we are told that no theory can be falsified unl theory is simultaneously corroborated. We must h hand before any can be superseded. 'Science' says any "refutations" leading the way' ([1970], p. 121) proliferation of theories is not merely beneficial, b in these circumstances no theory could be impug been established, so nothing would ever stimulate occasion and no good reason ever to invent a rival accepted; and it is difficult to understand what, in t scientist would feel obliged to try to explain, or in moved to invent new hypotheses. In that case the the abandonment of old theories would inhibit impossible of fulfilment.

Moreover, the history of science once again beli have always been proposed only when anomalies an in the old. Copernicus proposed the heliocentric because 'the mathematicians are so unsure of the Moon that they cannot even explain or observe seasonal year'. He complains, in the Prefatory let *De Revolutionibus Orbium Coelestium*, that Ptolema principles of explanation for the movement of di short, the old theory was incoherent. Harvey soug function of the heart because of the numerous co Galenic doctrine. Lavoisier proposed a new hypot and calcination because the observed facts of aug dicted the phlogiston theory (or, what is the same inconsistencies into it). Bohr (as Lakatos himself theory of the atom, because, according to classica atom, instead of radiating uniformly, should continu collapse into a neutralized nucleus. It is always the that stimulate research and prompt the proposal of

Further, Lakatos makes it a condition of accep predict new facts some of which must be corroborat predicted no new facts and was nevertheless accept the teeth of the available evidence.

A problem shift, Lakatos maintains, is progre hypotheses are empirically richer in content. But c on the orbit of Mars. His research programme pres what Lakatos calls a problem-shift. First he assume Then, when that showed a discrepancy of eight mi ovoid, produced by the introduction of an inverse he worked with an ellipse. But none of these succ new facts, for all the necessary empirical evidence was available from the start in Tycho's records. Nobody, however, would wish to deny that Kepler's problem-shift was progressive. Of course, if one asserts that the new fact so discovered was the elliptical shape of the orbit, one is saying only that Kepler's second law was thus established. But the law is a theory, not a piece of empirical evidence, and the claim now made for it is that it has been verified—a claim inadmissible for staunch Popperians.

As if aware that his theory, as so far developed, is not fully satisfactory, Lakatos adds yet another epicycle. The condition stated for acceptance of a new research programme is that it should be able to predict novel facts and that some of these should be corroborated. But now we are told that this depends largely on how we interpret factual novelty, and that 'the novelty of a factual proposition can frequently be seen only after a long period has elapsed'. (Lakatos [1970], p. 155.) In the final issue it is only when the 'hard core' of the *new* programme has 'become well corroborated' and qualifies as an 'observational' or interpretative theory that the facts which it has predicted are seen to be novel. Consequently, there are no crucial experiments which give instant testimony[1] of the failure (or 'degeneration') of one research programme and the superiority of another. It is only by hindsight and long after the event that experiments (like Michelson-Morley) come to be recognised (and, it would seem from Lakatos's account, falsely) as crucial. (Lakatos [1970], pp. 156 ff.)

In short, it is only in the light of a new hypothesis that *that hypothesis* can be seen to predict novel facts. It is only when it has become accepted as an interpretative theory that these facts can be held to have been corroborated. Thus it can only be *accepted* as progressive, and as requiring the rejection (or suppression) of its predecessor *after* it has been adopted as preferable. So the criterion for acceptance, and for falsification, is either specious or impossible of satisfaction.

5 In these circumstances, would it not be advisable to scrap the epicycles and transfer the centre of the theory from the 'empirical basis' to the consilience of hypotheses? Let us have done, once and for all, with Empiricism and all its remnants: 'basic' statements, novel facts, and unexplained processes of corroboration. Let us follow the Copernican example and shift the focus of our doctrine from observation statements to systematic construction of mutually consistent theories. We can then offer an account of scientific method and advance which includes all the features of Lakatos's methodology of research programmes, which agrees with historical facts, which heals the discrepancies and fills the blanks left by Popperism, and which remains internally consistent.

The passage in Lakatos that points the way to this Copernican revolution (not unlike Kant's) is the following:

> '. . . experiments do not simply overthrow theories, . . . no theory forbids a state of affairs specifiable in advance. It is not that we propose a theory and Nature may shout NO; rather, we propose a maze of theories, and Nature may shout INCONSISTENT. The problem is then *shifted* from

---

[1] Lakatos calls the immediate falsifying testimony of empirical evidence 'instant rationality' (the availability of which he denies), clearly indicating that, for him, rationality and empiricism are synonymous.

> the old problem of replacing a theory refuted by 'facts' to the new problem of how to resolve inconsistencies between closely associated theories.' ([1970], p. 130.)

To effect the revolution we must reexamine the source of all empirical knowledge, that is, perception itself, in order to decide the character and function of observation. When we do so we find that perceptual assurance is never the result of an immediate intuition (*Anschauung*), but is founded first upon a skilled activity structuring the confused contents of primitive sentience and ordering them in accordance with schemata. Some of the schemata are (or may be) innate, others are learned in the course of experience, but none are altogether exempt from modification, if, as the activity proceeds and experience accumulates, conflicts arise between schematic arrangements so that mutual adjustment of schemata is required in order to restore consistency. Perceptual assurance, therefore, is founded, secondly, upon the coalescence, without mutual conflict, of perceptual schemata. It is the result of the convergence and coherence of perceptual cues, the consilience of perceptual 'hypotheses'.[1] These schemata develop in the course of experience and become highly sophisticated. They do so when and as experience brings to light conflicts and contradictions in our habitual interpretations, which drive us to reflection and to the construction of explanatory hypotheses. In this way the common-sense awareness of the world gives rise to science and ordinary perception develops into scientific observation.[2]

We now have the basis of what Popper and Lakatos refer to as the 'unproblematic background of knowledge'. It is only relatively unproblematic, because as knowledge progresses new conflicts may arise at any level, calling for new efforts at reconciliation of schemata. But it is unproblematic so far as it is, for the present, free of such conflicts, or so far as those which do exist are of little or no obvious significance for the solution of current problems.

The contradictions which do arise and are troublesome both in theory and practice provide the *nisus* to research. Hypotheses which are proposed to resolve inconsistency are not invented arbitrarily, and do not spring from the heads of scientists like Athena from the head of Zeus. (Cf. Lakatos [1970], p. 133, n. 4.) They are developed out of prior (unsatisfactory or partially satisfactory) theories and are suggested by the structure of the facts as interpreted by those theories. This does not prejudice the facts in favour of outdated theories because it is precisely the discrepancies and inconsistencies of those theories in their interpretative and explanatory functioning that makes them unsatisfactory.

So Copernicus' hypothesis was taken over from Aristarchus, and its detail was worked out on the model provided by Ptolemy (for Copernicus needed epicycles as well to make this theory consistent). Kepler resolved the orbit of Mars by developing the earlier Aristotelian hypothesis of a circular orbit, through that of an ovoid (a circle distorted by an inversely revolving epicycle, using ideas borrowed from Ptolemy and Aristotle), to arrive finally at the ellipse of which the circle is a special case, and the formula for which emerged from his calculations to determine the ovoid. Harvey's hypothesis was prompted by earlier discoveries—the impenetrability of the septum, the valves in the veins, the lesser circulation through the lungs, all of them provided by his predecessors. It would be easy to multiply examples, but let these few serve.

---

[1] See my [1970], ch. viii; D. M. Vernon [1962], and E. H. Gombrich [1960]. Also J. S. Bruner [1957].

[2] I have tried to describe this process in more detail in my [1970], ch. ix.

It now becomes evident that scientists need to make no arbitrary, conventional decisions to determine what counts as fact, what experimental techniques to recognise, and the like. The interpretative theories develop with and are embedded in the unproblematic background of knowledge which has grown up through the centuries as science progressed. What counts as fact is what scientists recognise as sufficiently attested by experimental methods which have hitherto proved adequate and fruitful. These cannot be arbitrarily changed and may be modified only if the result is greater consistency. The interpretative theories are those hitherto found to be coherent, at least up to a point. Explanatory theories are those which, as their implications are developed, remove inconsistencies that have arisen in their predecessors. These elements of scientific interpretation and explanation are not separable, but constitute a single, though complex, system of schemata, the mutual consistency of which is the indispensable condition of ultimate intelligibility.

Next, we should note, that hypotheses in science *are* capable of 'verification', although the method is never mere enumerative induction. The word 'verification', however, must not be given too absolute a sense. No theory is *absolutely* true, so that it could never be called in question or modified at some later stage in the development of science. But very many theories are so well established that nobody would, at our present stage of knowledge, dream of doubting them. Harvey's is a case in point; and there is not much danger of a challenge to the heliocentric view of the solar system. This is not simply because such theories have not yet been falsified by negative instances. It is because the supporting evidence is systematically interrelated in such a way that the denial of any significant fact within it would bring whole edifices of science tumbling about our ears. Theories are established by the construction of such systematic bodies of evidence the process of which is the same quest for consistent and coherent interpretation that we have already noted. No better example can be given of the establishment of a theory in this way than the researches of Lavoisier, whose experiments built up, piecemeal and systematically, a structure of evidence the interconnected character of which was such that it could not be overthrown without rejecting all the results obtained by his immediate predecessors, Priestley's included.

Lakatos is right, therefore, to demand corroboration of a theory by a wide range of facts for its acceptance. But it must not only be rich in empirical content. Its empirical support must also be so inter-related that it constitutes a coherent system any part of which involves and requires all the rest. If, in the attempt to construct such a system, intractable contradictions or discrepancies arise, efforts are made to modify the theory in such a way that it will predict new experimental results that can be fitted together, both among themselves and with those previously obtained, so as to recover consistency and coherence. Here then we have the general principle of discovery and advance, involving both the refutation of old, and the establishment of new theories. The former is not effected (even after anomalies have appeared and doubts have been raised) until the latter is complete, because no theory, which has once proved valuable as a relatively coherent way of organising a body of facts, can be summarily abandoned, before another, which will do its job and do it better, has been devised and fully corroborated. And no theory would ever be adopted unless it did make some body of facts at least relatively intelligible.

Hence there emerges the problem-shift required by Lakatos. As the implications are developed of one theory, hitherto accepted or hopefully proposed, incoherencies arise; and that requires a modification of the hypothesis and the articulation of the revised form, in the attempt to remove the inconsistencies. If this is unsuccessful the problem-shift is not progressive. If it is successful it will result in a widening of the scope of the theory to cover both sides of the former inconsistency, and so will be progressive. Thus the first version of Bohr's atomic theory explained the hydrogen spectrum and resolved the contradiction in the classical interpretation of the Rutherford atom, but failed to account for the Pickering-Fowler ultraviolet series. Bohr's second version reconciled contradictions both within the former set of facts and between that and the latter set.

The dialectic of progress has four stages: (*a*) the construction of a theoretical system, (*b*) the emergence of contradictions in the course of its articulation, (*c*) the proposal of a new hypothesis, (*d*) the reconstruction of the system on the basis of the fresh hypothesis so as to remove the contradictions. Each stage is a comprehensive whole of empirical evidence organised by a system of theoretical schemata; and successive stages advance upon their predecessors as more comprehensive and more coherently integrated systems. The criterion of scientific acceptability is systematic coherence and that of scientific preference is comprehensiveness. (See my [1970], ch. xi.)

What now is to be said of the criterion of distinction between science and metaphysics? This is of primary importance only so long as metaphysics is held to be unscientific; but by the criterion of scientific acceptability that we have just offered this need not be the case. If metaphysics is an attempt to interpret our experience as a whole in a systematic and coherent way, it is scientific in character, and its distinction from natural science must be sought in some respect other than its scientific or non-scientific character. (Cf. my [1967].) The important demarcations are those between science and pseudo-science and between either of these and superstition. Science is systematic thinking that can maintain its consistency over wide areas of experience. Pseudo-science maintains only an appearance of system, which breaks down as soon as rigorous articulation is seriously attempted; and superstition is unsystematic and cannot maintain consistency in any considerable degree.

6 The example, given by Lakatos, of the development of the quantum theory fits the account of science which I have outlined above. The facts which supported the Rutherford model of the atom were interpreted in terms of the electro-dynamic theory of Lorentz and the classical theoretical concepts which had developed since Maxwell. So interpreted the model brought order and consistency into the explanation of numerous facts, including the way in which particles were deflected and X-rays scattered on impact with matter. But it involved a contradiction when used to account for the emission of radiation by the atom, on the assumption that a vibrating electron emitted radiation and continuously lost energy in the process. Bohr's theory, in its earliest version

(1913) resolved this contradiction and accounted for much of the detail of the hydrogen spectrum besides enabling him to calculate *a priori* the Rydberg constant. But in this form the theory came into conflict with the Pickering-Fowler ultraviolet series—a fact interpreted in terms of existing spectroscopic theory. Accordingly, a second version of the Bohr model was evolved in terms of ionised helium, which removed and explained the earlier discrepancy. But even this version of the theory fell short of accounting for all the facts. It had been based on the assumption of circular electron orbits and could not deal with elliptical trajectories which were demanded by other assumptions involved in the dynamics of the Rutherford-Bohr model (e.g. Coulomb forces governing the relations between the nucleus and its orbiting electrons). This defect was remedied by Wilson and Sommerfeld in 1916 and further comprehensiveness was attained concurrently with greater consistency of interpretative and explanatory hypotheses. Nevertheless, the whole theory combined classical and quantum concepts which were incompatible and was far from free of anomalies (e.g. its prediction of the doublets in the Balmer series and of X-ray spectra was exact, but put them in the wrong place). These final inconsistencies and discrepancies were resolved only after the 'old quantum theory' had given place to the new, as developed by de Broglie and Schrödinger.

In all this I have passed over what Lakatos calls the 'protective belt' of the hypothesis which forms the hard core of the research programme. This is the complex of auxiliary hypotheses which take the brunt of the 'refutations' or anomalies when they occur, and protect the main hypothesis from rejection in consequence of counter-instances. I have done so because this way of viewing the matter is not correct. The auxiliary hypotheses are all part of one complex system including the central 'core'. Anomalies are inconsistencies within this system, between various schemata that the over-all theory is attempting to combine. They do not overthrow the 'core' so long as adjustments can be made which enable research to continue and keep the problem-shift progressive. But when they become intractable, and if they multiply beyond the limits within which consistent progress can be made, a crisis occurs which demands more far-reaching revision of the system and is the source of a major revolution.

7 The example we have been examining is that of a minor cycle within a major revolution in the history of physics. Another such cycle in the history of astronomy was the progress from Copernicus to Newton. Periods of relative stability between major revolutions are periods of what Kuhn regards as 'normal' science. They are characterised by steady and relatively even progress of the articulation and application of a system of schemata, or concepts, under the aegis of which research proceeds. The development of the system invariably reveals minor conflicts and the process of removing them—Kuhn's 'puzzle-solving'—articulates the system further. As long as all goes smoothly the main concepts appear inviolable and anomalous cases, which cannot be removed (as long as they are isolated and have no obvious repercussions elsewhere) are ignored, shelved, or explained away by *ad hoc* devices. The condition for maintenance of the 'hard core' theories, is, as Lakatos asserts, continued successful progressive research within the system.

But these neglected anomalies are never without significance, and almost invariably prove in time to be symptoms of deeper inconsistencies, sometimes in underlying presuppositions that have been tacitly made and are never called in question or even brought to clear consciousness, until a crisis situation forces them upon the attention of scientists. Consequently, periods of such 'normal' and relatively straightforward progress are limited. The disregarded anomalies are bound to accumulate and the underlying conflicts that produce them are bound to have other and related consequences. When this accumulation of incoherencies reaches a critical point, as it did at the end of the last century in physics, something more is needed than mere mutual adjustment of accepted schemata. The whole system has to be revised, and major reorganisation is called for. The outcome is a scientific revolution and the establishment of a new conceptual structure.

Kuhn's distinction between normal and revolutionary science is, therefore, not unjust, but he errs in making it too sharp. He separates the two species of scientific progress as if they were two different kinds of science with nothing in common. Consequently, he fails to provide any explanation or logical basis for the transition from one so-called 'paradigm' to the next, and represents it simply as an inexplicable *gestalt* switch.[1] This is because he fails to recognise the dialectic running through the entire process, which is the same in both 'normal' and 'revolutionary' science. In the first, however, it affects only the minor constituent schemata of the over-all conceptual system; in the second it affects the conceptual scheme as a whole. In the first (normal science), the dialectical relation is between two versions of a subsidiary hypothesis. In the second (scientific revolutions) the dialectical relation is between complete systems.

In his response to his critics, ([1970]), Kuhn makes some significant and valuable clarifications and advances. He shows that the notion of 'normal' science is logically indispensable to that of revolutions, and he illustrates historically how the former may contribute to the precipitation of the latter. But he still falls short of explaining the dialectic operative in revolutionary science in such a way as to make its relation to 'puzzle-solving' in normal science fully intelligible. Toulmin, in his [1970], rightly points out the relative character of Kuhn's distinction, but does not (any more than Kuhn) reveal the dialectical process common to both phases of scientific advance. Lakatos comes nearer to providing what is needed, and I have made some attempt to do so more fully in a recent work. (Cf. my [1970], Part III.)

ERROL E. HARRIS
Northwestern University

---

[1] The analogy to a *gestalt* switch is not altogether misplaced. Something of the kind does occur in the course of discovery, when a scientist has been struggling in terms of one conceptual schema to reconcile conflicting interpretations of experimental facts, or to solve theoretical problems, which become intelligible only in terms of a new, or modified, schema. When the new pattern of interpretation dawns, as it often does quite suddenly, the scientist is prone to exclaim, with Archimedes, '*Eureka*', or with Kepler, '*O me ridiculum! perinde quasi . . . non posit esse* . . .'

REFERENCES

Blanshard, B. [1939]: *The Nature of Thought.*
Bosanquet, B. [1911]: *Logic.*
Bruner, J. S. [1957]: 'On Perceptual Readiness', *Psychological Review*, 64, pp. 123–152.

GOMBRICH, E. H. [1960]: *Art and Illusion.*
HANSON, N. R. [1958]: *Patterns of Discovery.*
HARRIS, E. E. [1967]: 'Method and Explanation in Metaphysics'; in R. E. Wood (*ed.*): *The Future of Metaphysics,* 1970, pp. 190–203.
HARRIS, E. E. [1970]: *Hypothesis and Perception. The Roots of Scientific Method.*
JOACHIM, H. H. [1948]: *Logical Studies.*
KUHN, T. [1962]: *The Structure of Scientific Revolutions.*
KUHN, T. [1970]: 'Reflections on My Critics', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge,* 1970, pp. 231–78.
LAKATOS, I. [1970]: 'Falsification and the Methodology of Scientific Research Programmes', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge,* 1970, pp. 91–195.
POPPER, K. R. [1935]: *Die Logik der Forschung.*
POPPER, K. R. [1959]: *The Logic of Scientific Discovery.*
TOULMIN, S. R. [1970]: 'Does the Distinction between Normal and Revolutionary Science hold Water?', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge,* 1970.
VERNON, D. M. [1962]: *The Psychology of Perception.*
WATKINS, J. W. N. [1970]: 'Against "Normal Science" ', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge,* 1970, pp. 25–37.

Brit. J. Phil. Sci. 23 (1972), 69–78 *Printed in Great Britain*

Review Article

BACK TO ARISTOTLE?

1 This book[1] has this in common with Kant's *Critique of Pure Reason*: it describes itself as describing a 'Copernican Revolution' (p. 15).

But whereas Kant's revolution was redescribed by Russell as a Ptolemaic counter-revolution, this one counter-revolves even further: it is an Aristotelian counter-revolution.

The central problem is the '*Problem of Induction.* How do we justify logically our entirely practically justified confidence in the continuance of certain, carefully distinguished, event "patterns"?' (p. 6). The author gleefully endorses Hume's criticism: 'it has frequently been pointed out, notably by Hume, that inductive evidence, the piling up of favourable instances, never can confirm a universal statement. . . . [The only other deductivist alternative would force us] into the absurd position of having to claim that the only significant evidence in science would be what would falsify a universal statement. The adoption of this horn of the dilemma as a methodological principle by Sir K. Popper . . . has allowed him to work out *as far as possible* a consistent deductivist philosophy of science. So we are impaled on the horns of a dilemma, either with the "problem" of induction, or with the absurdity of making out that all evidence in science is falsificatory.' (pp. 24 f.) ('Absurd because an *instance* falsifying a hypothesis depends upon the Uniformity of Nature equally with the confirmation of a hypothesis by an instance.' (p. 134)) Thus '[if] statements describing . . . [past] events . . . exhaust our knowledge, then no general statement [about] . . . so far unobserved events . . . can be known to be true' (p. 6). 'It seems clear . . . that it is the deductivist account that is at fault.' (p. 25)

The solution is a return to 'realism' (that is, essentialism). It is false that 'events are all that there is, and that events are independent one of another [the myth of events]' (p. 6). Rather, '[the] world consists of numerous, fairly permanent structures' (p. 10), 'causal mechanisms' (p. 11). 'The constitutions of such units are what Locke called "real essences", and their differentia constitute the natural kinds. The chemical analysis of a material, the genes inherited from a parent, the structure of a crystal, the nuclear and electronic constitution of an atom, are "real essences" . . ., and in them reside the powers of generation and production, through the operations of which the flux of events is generated.' (p. 10) For 'each event needs an explanation, and . . . gets it from the events which generate it under appropriate conditions and the mechanism by which this generation occurs' (p. 7). '[Theory] construction is primarily model building, in particular imagining paramorphs [see pp. 43–46]. . . . they introduce additional entities other than the given, provided it seems plausible to treat [them] . . . as a causal mechanism. . . . A scientific explanation of a process or pattern among phenomena is provided by a theory constructed in this way.' (pp. 46 f.)

---

[1] Harré, R. [1970]: *The Principles of Scientific Thinking*. London: Macmillan. £4.50. Pp. x+324.

The above quotations have been selected with some care, in an attempt to illustrate the author's philosophical position. Nevertheless, the reviewer is not sure that they will convey very much. The fact is that it is impossible to state clearly and succinctly, in the author's own words, the main theses of this undisciplined book. We will never illuminate if we don't speculate.

Briefly, Harré's argument is this (I think). Science is a rational activity; so if we look at what scientists do we should be able to discern rational principles at work, some of them, since scientists certainly achieve certainty, constructive ones. The main procedure of science is the explaining of observed regularities in terms of mechanisms. More precisely, a regularity is noticed amongst events (or states, or something). It is tested more thoroughly, and perhaps found to fail for certain ranges of values of the independent variables, or for certain subsets of individuals. Quite soon, however, its domain of application can be fixed (ch. 5). At this stage, as Hume showed, nothing is yet established or explained; and it is here that the constructive thinking begins. For unless the regularity is merely accidental, there must be some mechanism at the back that explains it. The scientist will therefore attempt to imagine a mechanism whose very essence or nature it is to generate, in suitable conditions, processes of the observed kind. To this end he will build models which simulate or ape, more or less exactly, the unknown essence of the responsible mechanism (ch. 2). Later, if the model is plausible, and works well, it may even be possible to verify that mechanisms like it exist behind the phenomena (ch. 3). In this way the regularity will be explained; the statement of regularity will come to be treated as almost irrefutable, as a law of nature (ch. 4).

The central contention of the book is that by these means *the problem of induction can be evaded, and scientific explanations established for certain.* The most straightforward strategy for rebutting this contention would be to say things that have been said several times before. But though, of course, it is possible that the author would be persuaded after one more repetition, it is not very promising. And, indeed, were it a matter of having to argue again Hume's arguments that no ampliative (content-increasing) inference whatever can be justified; or of having to offer once more the opinion that, although connections between events *cannot be known with the certainty of logic*, it is rank positivism to conclude that they do not exist; or of having to urge for the umpteenth time the undesirability of using essences; then, the reviewer feels, very little could be achieved. But as it happens it is possible, independently of any of these doctrines, to refute this central contention of Harré's; and it will be done in the next section.

The third section will look somewhat askance at the metaphysics revealed, in the later chapters of the book, at the core of Harré's epistemology. Harré—like many philosophers who have tried to solve the problem of change—finds the world populated ultimately by a swarm of unchanging entities. He rejects, however, with arguments that it will be instructive to scrutinise, a corpuscularian approach to these ultimate entities, and adumbrates instead a theory (consciously derived from that of Boscovič) of point atoms attracting and repelling one another by their very nature. It will be argued that this metaphysical essentialism is as unattractive as the epistemological essentialism built round it; and should likewise be rejected.

There being one section on epistemology, and one on metaphysics, it must not be supposed that the absence of one on logic (the fourth section merely lists some of the book's more serious errors) expresses the reviewer's satisfaction with what Harré has to say on this score. Quite the opposite is true; but Harré has so unsound an appreciation of the role of deductive logic that his polemic can perhaps be written off as based on a simple misunderstanding. Harré's opinion (derived from Toulmin) is that deductive logic is bankrupt as a fund of constructive rational principles (perhaps this is why it is frequently called 'received logic'), and that in consequence 'a richer system of rational principles' (p. 1) must be developed. Too true, if deductive logic pretended to provide a 'logic of discovery'; but it does not, and anyway there is no such thing. Logic's business is with criticism (and to a lesser extent with proof); not with discovery.

This is not of course a verbal dispute. Harré believes that scientists do use a 'logic of discovery'; that they do establish 'conceptual connections'; that they argue by analogy; even that they affirm the consequent (p. 23). Perhaps so. But because he takes a naturalistic attitude to science (strangely discordant with his violent assailment of positivism in ch. 1), he wishes to elevate, what are, logically speaking, aberrations (though perhaps heuristically useful) into 'principles of scientific thinking', into 'logic'. This can not contribute much to an understanding of (scientific) rationality.

As an example of the trouble we can get into, consider Harré's inconsequential criticisms of the deductive model of explanation (pp. 15–21). (1) The explanation by 'analogy' which he claims on page 16 not to be explanatory for a deductivist can easily be treated as an explanation-sketch, to be expanded by the addition of some universal law or laws (on the other hand, if the analogical explanation is held not to be in need of supplementation then the reviewer, for one, does not understand it). (2) In the treatment, on pages 17 f., of the condition that the explanations must have empirical content, Harré goes right off the rails in suggesting that this means that all predicates have been ostensively defined. It means that the explanans is testable, or confirmable. (3) The suggestion on page 18 that explanans of the deductivist kind cannot be established for certain is of course true, but hardly relevant, as the same is true of all others. (4) Nor are Harré's arguments against the symmetry of explanation and prediction very persuasive. He tells us that the 'absurdity of the alleged explanation/prediction symmetry is, or should be, immediately obvious [since] . . . predictions . . . [are] very rare in science, and explanation very common' (p. 18); but the symmetry is, as Harré acknowledges, supposed to be a logical matter, so the citing of historical or sociological evidence is quite beside the point (this is admitted on p. 20, but simply not answered). (5) For the same reason the remark that an evolutionary explanation could not be a prediction because 'the choice of . . . initial conditions depends upon our already knowing what was actually evolved' (p. 19) can be disregarded; though if true it would not say much for the explanation. (6) The accusation that 'in the deductivist theory, taken strictly, any further step beyond . . . the deduction of the phenomena from a law' is 'gratuitous'

(p. 21) would embarrass only the most positivistic of deductivists. (7) Last, it would indeed be 'absurd to say that the symptoms explained the subsequent syndrome' (because a universal law is needed); but it is quite as absurd to suppose that a prediction can be made from the symptoms without the mediation of a universal law (p. 21).

No further attempt will be made here to answer Harré's attack on deductivism. (Incidentally, deductivism is grotesquely mixed up with positivism in the author's discussions.) For Harré's whole position is simply swept away when we realise that even the best scientists, or rather, especially the best scientists, are always making mistakes; and that *deductive logic is our way of detecting these mistakes.*

2 It is with some diffidence that the reviewer ventures his interpretations of the central argument, and central contention, of Harré's book. Only something like what follows, I feel, enables us to disentangle some of Harré's pronouncements (especially the discussion of the 'Problem of Induction' on p. 248; though I rather doubt whether any interpretation could really clear this particular passage up); yet the thesis that I here attribute to Harré is nowhere stated, barely argued, and totally unfounded. It is this:

> (1) Lawlike regularities cannot be justified or explained by universal laws, because this leads to an infinite regress.
>
> (2) We can explain the *existence* of a regularity by the *existence* of a constant unchanging mechanism behind it. We can justify the statement that the regularity exists by *showing that there exists a mechanism behind it.* No universal laws are needed here.
>
> (3) We can explain the *nature* of a regularity by the *nature* of the responsible mechanism. We can justify the statement of the particular character of the regularity by *plumbing the nature of the responsible mechanism.* No universal laws are needed here.

That is, we can justify a statement of universal regularity without recourse to further such statements. For the *verification of existential statements* and the *divining of essences* are both non-inductive tasks. The problem of Hume is evaded.

The obvious objection to this is that Hume's argument shows a lot more than that a universal statement cannot be validly inferred from singular (or 'instantial') statements; it shows too, I should say, that no factual statement whatever can be established beyond all doubt; or that no content-increasing inference whatever can be justified. Thus statements of essence, since they are obtained by 'analogical inference', cannot be established as indubitable, *except as a matter of convention.* This, it seems to me, is how Harré has to see them, despite his disclaimer that realism 'accepts neither deductivism nor positivism nor conventionalism' (p. 1), and despite his plea (pp. 196–200) that taxonomic classification is not arbitrary.

But as I said before, it is not here that I wish to join battle. Rather I wish to disclose the extremely dubious part the 'unchanging mechanisms' play in this argument. Though he neither argues the case, nor seems to think that it needs argument (p. 248), we can perhaps accept Harré's neo-Parmenidean postulate that there is, lying behind the phenomenal world of change, a realm of unchanging entities. What we must not accept is the suggestion that by postulating such entities we can explain the existence of regularities wherever they occur. Harré of course says that regularities do not need to be explained at all; but what he means is that the explanation, being not causal (for 'enduring has no causes' (p. 248)), is trivial and does not lead to any further problems. Indeed, it does sound trivial to account for the existence of 'regularities of appearances . . . [as] the consequences of the endurance of internal structures' (p. 248); and one who denies the existence of permanent things altogether ('events are all that there is'), or even their role in generating or connecting events ('events are independent of one another'), does seem to be worse off. For by his insistence on universal statements he restricts himself to explaining regularities in terms of other regularities; and this allows of no 'distinction between laws and accidental generalizations' (p. 92). But the superiority of the Harré approach is in fact entirely spurious; for the triviality of explaining regularities by indicating the permanent entities in the background is quite illusory. Indeed, the explanation simply does not work; here too regularities can only be explained in terms of further regularities. For the whole point of positing permanent entities was to explain *how change is possible*; such entities cannot therefore *on their own* explain *why change does not occur*. (Harré may argue that he posits them for the other purpose; but then he has failed to solve the problem of change, which, incidentally, he nowhere mentions.) Nor will it be sufficient to introduce a statement asserting 'the stability of the conditions in which the events occur', in the hope that this stability can provide the needed permanence; for the conditions are always in some respects stable, even when the outcome is as chaotic a sequence as can be imagined. Thus in order to explain why there is a regularity we are obliged to formulate a universal hypothesis asserting that when the conditions are stable in the relevant respects then the regular sequence of events obtains. It is not, as Harré supposes, that the *existence* of the regularity can be explained by the *existence* of the mechanism; it can only be explained by universal laws.

This does not mean that explanations in terms of mechanisms are in some sense undesirable. Indeed, Harré does not 'intend anything specifically mechanical by the word "mechanisms". Clockwork is a mechanism, Faraday's strained space is a mechanism, electron quantum jumps [are] . . . a mechanism, and so on' (p. 36), so the opposite, if anything, is meant. But what is the point of *requiring* explanations to be in these terms? The answer is, of course, that the *particular character* ('intrinsic nature') of the mechanism is meant to explain the *particular character* of the regularity. And we cannot have an essentialist explanation unless we can find some 'thing' to carry the essence. That is why a mechanism is so important. But since Harré's attempt to avoid universality has truly broken down, the need for essentialist explanations disappears. (Harré does not believe in the possibility of *complete* explanations—those leaving nothing in need of explanation—, so essences solve no problems in this respect.)

Our argument so far, then, must lead us to conclude that, since regularities can only be explained by further regularities, essentialist explanations have no especially attractive properties, so that a restriction to them is unimaginative. In other words, the 'Copernican Revolution in the philosophy of science' is a failure.

3 In Harré's world essences exist, but they cannot be fully plumbed. This is expressed most clearly in his discussion of 'powers' in Chapter 10 (see especially pp. 270–274). Harré sets great store by this discussion, which he clearly considers to be of some importance. I regret to say that beyond repeating that all predicates are dispositional (if it even says that), it seems to me to be worth very little. Harré's 'proper analysis of the ascription of a power to a thing' as (p. 270)

> X has the power to do A = if X is subject to stimuli or conditions of an appropriate kind, then X will do A, *in virtue of its intrinsic nature*

shows, I fear, the emptiness of his essentialistic explanations. I agree of course that we should not be afraid of 'occult' powers; such positivistic fears need not concern us. But I disagree that they can eventually be reduced to essences or natures. For our laws of nature are expressed in terms of *relations*. Because the world is a world of interactions it is impossible to explain it, except entirely metaphysically, in terms of the 'natures' of its constituent parts. In particular it is impossible in principle to make an absolute or logical distinction between a thing's nature and its environment (the stimuli or conditions it is subject to); a fact which should be clear to anyone acquainted with discussions of 'intelligence' in educational circles. (Though our concern here is with physics, this example is significant; for, since Aristotle, biology has always been peculiarly prone to essentialist excesses, and Harré himself is clearly one of those inspired by the essentialistic character of Mendelian genetics.) We are always obliged to make use of universal laws. Consequently Harré's analysis of power ascriptions is objectionable; for it dogmatically assigns everything unknown about why X does A to X's nature, rather than to its surroundings. But this is a mistake.

A fundamental difficulty for the mechanical theory of the world was the problem of the explanation of its fundamental interactions, first of all impact, and later gravitation. Descartes supposed that the behaviour of bodies in impact could be explained by the nature of matter, its extension. But as Leibniz showed, extension does *not* explain how matter can *act*, nor therefore how it behaves. And though we can incorporate some principle of activity into the essence of matter, we do not by this means explain the behaviour of matter. For in any impact a particle is, it seems, as much acted upon as acting, so that its behaviour is dictated, at least in part, not by its intrinsic nature but by other particles. For the same reason, it is unacceptable to describe matter as essentially something which, in certain conditions, behaves in a certain way. The trouble is that since the 'conditions' are produced by the behaviour of other matter, any explanation reduced to these terms is more or less circular (or involves some other form of vicious infinite regress). Some alleviation can seemingly be obtained by treating as fundamental not the elementary interacting particles (corpuscles, or dynamic atoms, or whatever), but whole (isolated) systems; the components of a system behave as they do because it is the essence of such a system that they should do so. But because no system is completely isolated, the interaction of systems has to be explained too; so that the respite provided by this dodge is only temporary, and therefore nil.

We see therefore that there really is no absolute distinction to be made between a thing's potentialities and the conditions under which these potentialities are actualised. Harré's complicated discussion of the difference between circumstances and conditions (pp. 271–3), and his claim that we can distinguish between a thing and its surroundings seem to me to be mistaken and to be based on the fact that we often make such a distinction in practice. But this is not the point at issue. It is simply not enough to assert that the distinction can be made (pp. 107 f.), when there are arguments to show that it cannot. But there is no possibility of arguing this point here in more detail. For I wish to turn to a brief discussion of Harré's revival of some of Boscovič's ideas on dynamic atoms; or, rather, since I have little to offer here, with his attack on corpuscular atomism. The argument (Boscovič, *A Theory of Natural Philosophy*, sections 17 ff.) is quite well known, and is interesting because it indicates that 'models' or 'mechanisms' must be treated with care even to establish *consistency*. It hinges on the fact that corpuscular atoms, being absolutely simple, must be absolutely hard, allowing of no deformation; consequently an impact is an instantaneous (that is, discontinuous) action. But this, according to Newton's laws, involves infinite forces. Consequently no such impacts can exist. But Harré appears to read in here something which escapes me, for he says that the 'argument depends upon accepting that infinite forces are impossible or that all action is continuous. The impossibility of infinite forces will scarcely be disputed, so the argument turns on only one condition, the Law of Continuity, as Boscovich calls it . . ., [which] "consists in the idea that . . . any quantity in passing from one magnitude to another, must pass through all intermediate magnitudes of the same class." ' (p. 287) Now no doubt a 'proof' of the law of continuity might be interesting, but it is hardly needed, as far as I can see, if infinite forces are impossible (unless we do not wish to rely on Newton's law). And this is just as well, since Boscovič's metaphysical 'proof', especially in Harré's formulation, is patently, even trivially, invalid.

Part of the trouble comes from the definition of continuity. Harré tells us that (p. 288) 'Boscovič, anticipating Dedekind, notices that to say that in a continuous sequence any two points are contiguous will not do, since if the sequence is genuinely continuous, there must be infinitely many points *between any* two points. His condition for continuity is [that any point] . . . can be regarded *either* as the last point of the lower segment [that it determines], in which case the upper segment has no first point, *or* as the first point of the upper segment, in which case the lower segment has no last point' (p. 288). Consequently, suppose 'in some continuous sequence of times there should be discontinuous action, [of] . . . measure a in the lower segment in the sequence of instants, and b in the upper . . . [where $a \neq b$]. . . . Q, the point of discontinuity of the action, . . . has to be able to be considered either as . . . the first member of the upper segment, in which case the action has measure b at Q, *or* . . . as the last member of the lower segment, in which case the action has measure a at Q. So, either at Q the action has measures of both a and b, which is a contradiction . . . or . . . no measure [how?] . . . which is contrary to hypothesis' (pp. 288 f.). Harré concludes that we need only show that time is continuous (assumed by Boscovič) to complete the proof; and he proceeds to attempt to prove this 'through what Kant called *Phoronomy*, the general theory of kinematics' (p. 289).

It is difficult to know where to start; let us note briefly however that Harré's 'proof' (pp. 290 f.) that we cannot have discontinuous time along with continuous space is not of much interest, since it *assumes* that infinite forces do not exist (and so it is not 'phoronomic'). But this is, given Newton's theory, almost the very thing at issue—the law of continuity of action. (Nor is the suggestion that 'discontinuous time is not really coherent' (p. 291) very convincingly argued; for Harré seems to depend entirely on a confusion between continuous *functions* and continuous *sets*—or continua.) And in so far as the law of continuity extends beyond action, the proof is downright ridiculous. For if the argument Harré gives us is valid, *there is no such thing as a discontinuous function of a continuous variable.* This point is so obvious, once seen, that I need scarcely labour it. It suggests at once that something must be wrong with the definition (anyway at most a density condition), as indeed there is; and though we can perhaps forgive Boscovič, we can hardly applaud Harré's repetition of the error, especially as he mentions Dedekind. He should at the very least have told us that, given a modern definition of continuity, the proof he offers fails to get off the ground. Moreover, it is, I think, at least possible that Boscovič would not have acknowledged the definition of continuity ascribed to him by Harré; for it clearly refers to open intervals, whereas Boscovič, in section 51 of his book says that 'there cannot be a finite line without a first and last point', and in section 48, with reference to Aristotle, more or less the same. That is, he takes all intervals to be closed. I have naturally no wish to defend this erroneous view.

Despite all this, we must admit that the hypothesis of corpuscular atoms is seriously embarrassed by Boscovič's penetrating criticism. But of course this is not to establish Boscovič's ingenious rival theory of dynamic atoms acting on one another at a distance. It is this theory, trimmed out with a good deal of talk about powers, that Harré presents briefly in the last chapter (has he forgotten his assertion on page 74 that '[by] and large the sciences do not recognize, and never have recognized, action at a distance'?). Apparently he wishes to 'offer . . . an *a priori* proof that the fundamental entities . . . must be centres of mutual influence, and that the degree to which they mutually repel and attract are different functions of distance' (p. 305). Now of course no such proof is possible, and it is a pity that Harré, if he wants to present us with a metaphysical theory, should try to do it in this aprioristic way. For it leads him to suggest that the setup he is offering is the 'simplest possible form of *field*' (p. 305). Maybe. But it differs from most classical fields in a most important way; namely, it is in an obvious sense discontinuous. For in order to explain elasticity, Boscovič had to postulate that the mutual interactions between point atoms become infinite at vanishing distance (note that this is an infinite potential rather than an infinite force). Consequently, only finitely many atoms occur within any given volume, and we do not have forces emanating from every point. Now Harré tells us nothing about this, doubtless because there is a grave difficulty in giving any intelligible essentialist interpretation of continuous fields. But this only makes more transparent the futility of essentialism at this level. And if at the ultimate level there are no essences, what hope is there higher up?

4 The points on which Harré has been criticised above have been selected from a very large number of candidates. The book is teeming with mistakes and misinterpretations (not to mention misprints). Here are some examples.

(1) Chapter 6 is devoted to the problem of whether 'the notion of the probability of an event [can] be extended, or developed by analogy, to provide an

account of the degree of confirmation of a hypothesis' (p. 157). Harré criticises with some severity the position taken by Carnap, but neglects to point out that almost all the valid criticisms that he gives (pp. 168-75) are due to Popper, who is himself the target of a somewhat ill-directed attack (p. 175): 'he [Popper] tries to define his extreme values of confirmation and disconfirmation as follows

$$-1 = C(\text{not -}h, h) < C(h, e) < C(h, h) = 1.\text{'}$$

A glance at *The Logic of Scientific Discovery*, page 400, desideratum (ii), would show that the two '$<$' 's and the second '$=$' should all be '$<$'; thus for Harré to say that 'Popper does not clearly disentangle the notion of "necessary truth" from "fully confirmed by evidence" ' (p. 175), and that 'Popper's third [*sic*] desideratum, which requires, *inter alia*, that all empirical hypotheses shall have the same range of confirmation values, must be rejected' (p. 176) is extraordinary. As for the remark that 'Popper fails to distinguish confirmation from failure to falsify', I can only ask: Did Maxwell fail to distinguish light from electromagnetism? Did Einstein fail to distinguish inertial from gravitational mass? (2) For other scarcely more admirable technical performances the reader is referred to Harré's 'numerical measure of the degree of belief we should have in a hypothesis' on pages 150 f. (no criticism necessary) and to the 'measure of natural necessity' on pages 114 f. (I need only note that $\mathcal{N}_k$ appears to be able to exceed 1; but this is a bit obscure, since it is not made clear what is the '*k*th possibility', nor why $\mathcal{N}_k$ is not symmetric in *k*). (3) I might note also the extremely messy discussion of pages 143–6, which is rendered almost unreadable by, amongst other things, the continued printing of '1' for '*l*'. (4) Logicians will not, I think, be impressed by the 'main conceptual innovation' of Harré's theory of confirmation (p. 147); for it turns out to be nothing but a bounded universal quantifier. (5) 'The conditions for a set of objects to constitute a pre-dynamical linear spatial sequence' offered by Harré on pages 237 f. are hardly more satisfactory from a formal point of view. For one thing Harré fails to remind us (what he had already noted on p. 183) that '*aSb*' means that '*a* and *b* are spatially separated'. No doubt it is clear. But also clear is the fact that the axioms for *B* ('betweenness') are so weak as to allow many interpretations, such as 'beyond' or 'beneath' (actually they are *inconsistent*, strictly speaking, as can be seen by letting $a = b = c$ in (2); but I suppose that to be taken care of). Nor can (4) be derived from (1), which in fact has the form of a definition. And Harré's discussion of 'whether the conditions . . . lay down . . . desiderata with which only spaces agree' (pp. 240–2) is almost too inadequate to be true. (6) The Principle of Epoch Indifference, ' "There are no physical consequences of existing at one time rather than another" . . . [,a] principle [upon which] ultimately hangs most of the rationale of the account of scientific thinking I have been giving' (p. 246) seems to be in direct conflict with the second law of thermodynamics; and certainly with some of Boltzmann's theories. (7) Harré embraces a verifiability theory of truth; see especially pp. 148 and 188. (8) There are a number of defects in Harré's system of references. First is the fact that no dates are attached to journal articles. Second is the obscurity of some of the references, such as the one to Poincaré on page 47 (no book cited) and the one to Reichenbach on page 239. Third, one might note a couple of obvious errors: reference 147 gives the wrong journal, reference 203 the wrong name. (9) Worse by far is Harré's irritating habit of scattering vague cross-references throughout the book. He is always going to 'return to this question', but he seldom says where. A bad example is the 'question of negated predicates' referred forward from page 23; it eventually turns up on page 120, but there is little discussion. Perhaps we must proceed to Chapter 7? Of another sort are examples where the cross-references simply fail altogether: the forward reference on page 129 to the 'bibliography to Chapter 7', or that on page 170 to the 'bibliography on "consilience"'. (10) To conclude, I will mention only the bad organisation, sadly turgid style, and less than perfectly intelligible summaries, which combine to make the book unpleasantly hard going. It is ironical that a book obviously directed against the dryness and dreariness of so much contemporary philosophy of science should itself be so unattractive, and, I fear, unrewarding.

DAVID MILLER
University of Warwick and
Minnesota Center for Philosophy of Science

Brit. J. Phil. Sci. 23 (1972), 79–84 *Printed in Great Britain*

Reviews

BUCHDAHL, Gerd [1969]: *Metaphysics and the Philosophy of Science.* Oxford: Basil Blackwell. £5.25. Pp. xii+714.

Up to the time of Hegel, philosophers were well acquainted with the science of their day. Some of them, like Descartes and Leibniz, were themselves scientists of consequence. But for a hundred years after Kant, only a minority of metaphysicians had any knowledge of, or indeed respect for, scientific findings. This was due in part to the immense growth of man's knowledge of nature which had the paradoxical effect of fragmenting knowledge and seemed to offer a choice between specialisation and dilettantism. It was partly due also to the inward looking tendencies of idealism. Idealist philosophers were deluded into thinking that mind and nature were so related that the nature of knowledge could be understood without attending to its most successful instance. That this has now been remedied is due in large part to the work of Russell, the logical positivists and Karl Popper. It is indeed still possible for a philosopher to avert his eyes from science but it is more difficult than it was. The rise of psychology and, more recently, of linguistics have closed the last avenues to successful dilettantism in philosophy.

Historians of philosophy however have been slow to recognise the reintegration of science and philosophy and have, in general, written of the classical philosophers of the seventeenth and eighteenth centuries as if they were as innocent of scientific knowledge and influences as Hegel, Bradley or McTaggart. Buchdahl's principal achievement in this book is to re-write the history of philosophy from Descartes to Kant so that we no longer have any excuse for these antiquated attitudes. He brings a knowledge of contemporary science and of its historical development to an illuminating re-description of the more important figures of the classical period of European philosophy.

The details of his account are, of course, open to dispute. Any interpretation of the philosophers of the past must be conjectural and incapable of being finally established. But Buchdahl's account has the advantage and authority of being more broadly based than most alternative versions and gives in important respects a richer and more unified picture of the period than we find elsewhere. I found his treatment of Locke and Leibniz, in particular, especially instructive. I should have liked to see his methods applied more widely to include Hobbes, Gassendi and Spinoza. They are not mentioned at all except for casual passing references. But this is perhaps an unreasonable criticism of a book which exceeds 700 pages. Kant figures, rather conventionally, as the hero of the book and as the great reconciler of his predecessors' somewhat wayward insights. But the chapter on Kant is a book in itself of over 80 thousand words. And this is, as Buchdahl emphasises, only a selective account of the Kantian philosophy 'from the viewpoint of the student interested in the philosophy of science'.

The book is an original and substantial contribution to the history of philosophy, set in a neglected but illuminating framework. As we might expect in a book of this size and scope, it is not entirely easy reading. But there is no uncalled for difficulty in the treatment. The two least successful chapters are the first and second which are not historical but sketch the philosophical background against which Buchdahl sets his critical history. I found these—and in particular, Chapter II, 'Laws of Nature and the Propositional Link'—difficult to understand. Chapter II is a very condensed critical survey of the different senses of 'necessity' applicable to propositions. (Buchdahl calls it a 'natural history' of the notion of necessity.) But this is too large a subject to be treated satisfactorily in 60 pages or so, raising, as it does, topics as varied and complex as the nature of the *a priori*, the relation between sentences, statements and propositions, truth, laws of nature and explanation, to mention only the most important. I found this chapter hard to follow; however I did not find that my lack of understanding prevented me from profiting from the succeeding historical sections (I mention this as Buchdahl claims that what he calls the 'propositional link' between subject and predicate is a topic which will 'serve as the thread guiding us through the labyrinth of argument'. This particular thread seemed to be both confusing and unnecessary.) But the book is, notwithstanding, an important and original contribution to the history of ideas.

D. J. O'CONNOR
Exeter University

BUTTS, Robert E. and DAVIS, John W., (*eds.*) [1970]: *The Methodological Heritage of Newton.* Oxford: Basil Blackwell. £1.75. Pp. 170.

Three of the papers in this collection were delivered at a conference at the University of Western Ontario in 1967; one was to have been delivered at this conference and the remaining three were prepared by invitation. If the adjective 'methodological' be taken to mean 'related to the determination of the truth of a scientific proposition' (but excluding techniques, such as quantitative analysis in chemistry) then four of the seven papers seem to me to be concerned with Newton's methodology and two with his metaphysics; the seventh paper, by Gerd Buchdahl, 'Gravity and Intelligibility: Newton to Kant' is rather logical in nature since it is concerned with the intellectual status of Newton's law of gravitation. Buchdahl rightly points out that considered as a causeless cause of phenomena (that is, in an action at a distance formulation) gravitation was as unintelligible to Newton as to his harshest critics; Buchdahl is not here concerned (as is F. E. L. Priestley in his contribution on 'The Clarke-Leibniz Controversy') to discuss the arguments by which Newton (or Clarke) sought to render gravitation intelligible by providing an explicans for it. Instead, he considers the moves by which Locke, Hume and Kant in particular attempted to make intelligible and therefore acceptable what had not been so to Newton himself, without extending the causal chain. All three seem a little uncomfortably close to Alexander Pope's well-known maxim.

In Priestley's essay on a well-worked theme the most interesting points (to me) are his convincing rejection of the Cohen-Koyré suggestion that Leibniz was *tamquamised* (by an aberrant printing of the 1706 Latin *Opticks*) into his criticism of Newton's identification of space with God's sensorium, and his analysis of Newton's discourse on God in the General Scholium. He remarks (p. 36) that "the real debate . . . about the sets of assumptions underlying the positions of each " never took place. John W. Davis, in 'Berkeley, Newton and Space', contends that Berkeley's writings before *Siris* [1744] exhibit his reaction to the first edition of the *Principia* only, and that therefore the *Queries* in *Opticks* had no important bearing on his thought; similarly, Newton was in all probability unaware of Berkeley's criticisms.

Turning now to the methodological papers, only one, L. L. Laudan's 'Thomas Reid and the Newtonian Turn of British Methodological Thought', is directly concerned with the heritage from Newton. Pointing out that the three British empiricists Locke, Berkeley and Hume were "surprisingly un-Newtonian when it comes to questions of scientific method", Laudan's claim for Reid (1710–96) is that he was "the first major British philosopher to take Newton's opinions on induction, causality, and hypotheses seriously" (p. 106). In short Reid was a thorough-going inductivist who erected Newton's dicta into what Feyerabend might call a theology. But Reid could hardly be called a subtle or imaginative thinker. Whewell's analysis of one gospel in the Newtonian Bible, the *Rules of Reasoning in Philosophy* (inserted, curiously enough, after the reader has perused two-thirds of the *Principia*) forms the subject of a very interesting essay by Robert E. Butts, whose core is perhaps best represented by these two sentences (p. 146):

> Both Whewell and Newton had an unwavering confidence in the powers of induction, but each had a different view of induction. For Newton (at least as Whewell read him) induction is collection of laws from phenomena; for Whewell induction only inessentially involves collection of data, what it involves essentially is the imposition of a new idea on the data.

But it is also worth examining his arguments that the *Rules* may be viewed not merely as inference-rules (as Whewell saw them) but as discovery-rules.

The book opens with a typically swashbuckling but disappointing paper from my much lamented friend N. R. Hanson, to whose memory it is dedicated. Formally complex, Hanson's exposition seems to reduce to the simple and not very original claim that whatever Newton's uses of the word "hypothesis" in legitimate scientific contexts, in *Hypotheses non fingo* he employed it as a pejorative synonym for "unverifiable speculation". Like Hanson, Paul K. Feyerabend in the most radical and exciting of all these essays ("Classical Empiricism") finds historians lacking in critical sense (see particularly p. 162, n. 9), though not in the same way. The body of Feyerabend's paper contains two chief points: (1) Newtonian inductivism (which is *not* to be identified with Baconian inductivism as Laudan rightly emphasises, pp. 119–23) is a self-confirming system comparable to Protestant theology, which may therefore be characterised as a "party line"; it protects itself from disconfirmation both by failing to specify what evidence or arguments would count against it and by formulating *ad hoc* hypotheses to accommodate known difficulties. (2) Newton was *in fact* wrong in his fundamental ideas about light, and his rejection of the wave-theory was baseless. Now although there is a clear relation between these two contentions—indeed, the second might be considered as a specific illustration of the epistemological hollowness to which the first draws attention—the second of Feyerabend's onslaughts on classical empiricist science is necessarily to some extent independent of the first, and to that extent partially open to a charge of unhistoricity. For example, Feyerabend appears to suggest (p. 168) that the ability of a transverse wave-theory to contain the "three fundamental properties of light . . . and provide means for calculating the others" was (or could have been?) known to Newton when he rejected that theory. I do not believe this to be true. But more seriously —and without disputing Feyerbend's interpretation of Newton's fundamental allegiance—only a very single-minded reader of Newton's *Opticks*, *Lectiones opticae* and correspondence could conclude that the chief object of all this labour was the justification of a material or emission theory of light; or that because this justification was later falsified all this labour was devoted to a vain, circular argument. I find Feyerabend's analysis immensely stimulating and the final twist to his essay is too amusing to give away in a review. One should enjoy it in the original. Yet I wonder whether Feyerabend's line may not induce a revival of positivism? Can any theoretical system escape the criticisms which he levels against both Newtonian physics and the quantum physics of the Copenhagen school? Though scientific theories are not born with psychological assurance (save perhaps in the minds of their creators) and though the dogmatic confidence of an *n*th generation may itself be the measure of a hard passage through the fire (easily forgotten except by historians) of course it is proper to distinguish, as Feyerabend does, between the psychological authority of a theoretical system and its actual logical and contingent validity. Yet, while admitting the truth of Butts's comment (pp. 132–3) that "Today hardly anyone [among the philosophers interested in methodology] is seeking methods that someone will actually apply in doing or in teaching science" the practising scientist feels a need to articulate some view of his work; should he be convinced that all theoretical systems are in some ultimate sense vacuous (a favourite word in this book)—even though obviously within a limited area useful—may he not then retreat to the rather dull view that "science consists of pointer-readings"?

Finally, though it is perhaps idle to comment on the title of a book, I would suggest that the methodological heritage of Newton is a much vaster and indeed more historical subject than is envisaged in this collection of essays, with the exception of Feyerabend's sweeping and radical analysis. To explore it thoroughly would necessitate traversing a great part of eighteenth- and nineteenth-century science, as may be seen from such studies as I. B. Cohen's *Franklin and Newton*, A. Thackray's *Atoms and Powers* and C. Truesdell's publications on the development of rational fluid mechanics (anything like a general history of classical mechanics remains to be written). "Methodology" is a word, I suppose, appropriate to philosophy; but the history of Newton's methodology can hardly be considered simply as an aspect of the history of philosophy—a point which is, of course, well illustrated by the double-edged nature of Feyerabend's criticism of classical empiricism.

A. RUPERT HALL
Imperial College London

ERWIN, Edward [1970]: *The Concept of Meaninglessness*. Baltimore and London: The Johns Hopkins Press. $6.50. Pp. 164.

One wonders if this is the forerunner of a series of books on such concepts as mindlessness, illogicality, obscurity, and non-existence. The author seems to regard the concept of meaninglessness as pretty central within modern philosophy, though we are not told why. He accepts what he calls the 'moderate skeptical' thesis that all extant criteria for meaninglessness are defective, but rejects the 'radical skeptical' thesis that any such criterion is bound to be defective. His aim is to rebut this thesis by supplying a satisfactory concept himself. He hardly discusses *why* the idea of meaninglessness was introduced into philosophy in the first place (though he does allude briefly to Einstein in connection with Bridgman's operationalism and Schlick's verificationism); nor does he consider whether there is still any need for it. His attitude seems to be: the concept may be in poor shape but it is *there*, so we must do something about it.

He disposes of operationalism in twelve pages and of verificationism in twenty. Some of his criticisms are brisk and workmanlike. He claims that an operational definition will, in general, be deficient unless qualified by some sort of *ceteris paribus* clause (if we define intelligence in terms of I.Q. scores, we should exclude scores that are unduly low because the subject's pencil broke, or unduly high because the subject had mugged up the answers beforehand); but how, he asks (with acknowledgements to Peter Achinstein and Rudolf Carnap), could such a *ceteris paribus* clause itself be operationally defined?

But some of his criticisms are *too* brisk. As a knock-down objection to any confirmability criterion of meaning he offers the following argument:—Any such criterion will be far too narrow unless it allows statements to be meaningful that are *indirectly* confirmable; but *any statement whatever* will be indirectly confirmable. For we can 'imagine a computer ... constructed ... so that it makes only true assertions. Further, every assertion made by the computer has been independently checked and found to be true.... Now, select any statement whatsoever (from among those not known to be true, false, or meaningless) and let it be asserted by the computer. The mere fact that the statement had been made by the computer would provide confirming evidence for the statement' (p. 37).

But how can we 'let' any arbitrary statement be asserted by a computer that makes *only true statements*?

The author has a disquieting habit of first quoting a view to be criticised and then sliding into an elaboration of it which he proceeds to belabour with arguments which would not disturb the originally quoted view. For instance: elaborating on a quotation from Bridgman, he declares that, for Bridgman, 'a concept would be meaningless if it *could not* be defined operationally' (p. 5, my italics). But the quotation justified no more than, '... is meaningless until it has been defined operationally'. (There is some slight justification for Erwin's elaboration on p. 28 of *The Logic of Modern Physics*.) Again: he quotes a passage from Popper which concludes, '... the expression "*a* is an element of the class *a*" turns out to be meaningless in one language [i.e. Russell's] but meaningful in another [e.g. Zermelo's]; and this establishes that a proof that an expression is meaningless in some languages must not be mistaken for a proof of intrinsic meaninglessness'. Two pages later, without any further evidence being adduced, Erwin alleges that Popper was claiming 'that *any* meaningless statement may be true [!] in some alternative language' (Erwin's italics).

The brand of alleged meaninglessness with which the author is mainly concerned is that involved in type or category mistakes. 'Virtue is square' is his chief example. He reaches the following conclusions. (1) The predicate 'meaningless' should be applied to *statements*, for two reasons: first, no *sentence* is true or false; second, *any* string of words *can* be used to make a true or false statement in some language or other. (2) A contradictory statement is false *and* meaningless. (3) A statement that involves a category mistake, such as 'Virtue is square', is very like a contradictory statement: it is false and meaningless, and its negation ('Virtue is *not* square') is necessarily true.

As to (1): anyone is free to choose his own terminological conventions, but this one seems awkward. To say that *x* is meaningless is to say, one would have thought, that *x* fails to be, or to express, a statement. On Erwin's terminology this would become: '*x* is a statement which fails to be, or to express, a statement.' Erwin himself is driven to such awkward locutions as, 'a category mistake is a particular *statement*' (p. 111).

As to (2): this makes nonsense of *reductio ad absurdum* proofs. If the assumption that there is only a finite number of prime numbers is *meaningless*, how did Euclid succeed in deriving a contradiction from it?

As to (3): although 'Virtue is not square' is said by Erwin to be *a priori* true, he adds that it is not analytic. Rather perversely, he raises the question whether there are false synthetic *a priori* statements, but does not ask whether there are *true* ones. Had he done so he would presumably have had to reply that there are and that 'Virtue is not square' is one of them.

It may seem an odd result that the positivist idea of meaninglessness should lead to a rehabilitation of synthetic *a priori* metaphysics (though admittedly of an unexciting brand). But when a philosophical idea is plucked from its historical problem background and played about with in what seems, to this reviewer, to be an essentially aimless way, odd results are rather to be expected.

The book includes a chapter on metaphors. The metaphor-industry has been expanding rapidly since it was launched in 1954 by Max Black's paper on metaphors to the Aristotelian Society. But what does this industry produce? The upshot of the author's lengthy discussion is that a metaphor should not be taken literally: 'Man is a wolf' does not mean that man is a wolf; it means, roughly, 'Man is like a wolf (in certain respects)' (p. 122).

J. W. N. WATKINS
London School of Economics

Brit. J. Phil. Sci. 23 (1972), 85–88 *Printed in Great Britain*

Recent Publications

BOOKS RECEIVED

Achinstein, Peter [1971]: *Law and Explanation: An Essay in the Philosophy of Science.* Oxford: Oxford University Press. £2.75. Pp. xii+168.

Anscombe, G. E. M. [1971]: *Causality and Determination: An Inaugural Lecture.* Cambridge: Cambridge University Press. 30p. 95 cents. Pp. 30.

Bennett, Jonathan [1971]: *Locke, Berkeley, Hume: Central Themes.* Oxford: Oxford University Press. £3.50 cloth; £1.40 paperback. Pp. x+361.

Berelson, Bernard [1971]: *Content Analysis in Communication Research.* New York: Hafner Publishing Co.; London: Collier-Macmillan. £5.50. Pp. 220.

Binkely, R., Bronaugh, R. and Marras, A. (*eds.*) [1971]: *Agent, Action and Reason.* Oxford: Basil Blackwell. £2.50. Pp. v+202.

Blumenberg, Hans [1971]: *Pseudoplatonismen in der Naturwissenschaft der frühen Neuzeit.* Mainz: Akademie der Wissenschaften und der Literatur. Pp. 38.

Bradley, J. [1971]: *Mach's Philosophy of Science.* London: The Athlone Press. £4.00. Pp. xii+226.

Buckley, Michael J. [1971]: *Motion and Motion's God.* Princeton: Princeton University Press. London: Oxford University Press. £4.75. Pp. 287.

Cairns-Smith, A. G. [1971]: *The Life Puzzle: On Crystals and Organisms and on the Possibility of a Crystal as an Ancestor.* Edinburgh: Oliver and Boyd. £1.95. Pp. viii+165.

Čapek Milič [1971]: *Bergson and Modern Physics:* (Boston Studies in the Philosophy of Science, 7). Dordrecht: D. Reidel Publishing Co. D.fl. 70. Pp. xv+414.

Carovillano, Robert L. and Skehan, S. J. (*eds.*) [1971]: *Science and the Future of Man.* Cambridge, Mass. and London: MIT Press. $10.00. Pp. xiii+196.

Castel, Helene [1971]: *World Development: An Introductory Reader.* New York: The Macmillan Co.; London: Collier-Macmillan Limited. 90p. paperback. Pp. xvii+296.

Čížek, F. and Hodaňová, D. [1971]: *Evolution als Selbstregulation.* Jena: VEB Gustav Fischer. Pp. 316.

Cohen, Robert S. and Seeger, Raymond J. (*eds.*) [1970]: *Ernst Mach. Physicist and Philosopher* (Boston Studies in the Philosophy of Science, 6). Dordrecht: D. Reidel Publishing Company. D.fl. 38. Pp. viii+295.

Collins, James [1971]: *Descartes' Philosophy of Nature* (American Philosophical Quarterly Monograph Series, (*ed.*) N. Rescher). Oxford: Basil Blackwell. £1.75; $6.00. Pp. viii+99.

Cornman, James W. [1971]: *Materialism and Sensations.* New Haven and London: Yale University Press. £4.75. Pp. 352.

Crombie, A. C. and Hoskin, M. A. (*eds.*) [1970]: *History of Science*, 9. Cambridge: W. Heffer & Sons Ltd. £2.50. Pp. 147.

Crosson, Frederick J. [1971]: *Human and Artificial Intelligence.* New York: Appleton-Century-Crofts. $4.40 paperback. Pp. vi+267.

de Bono, Edward [1971]: *Technology Today.* London: Routledge & Kegan Paul. £1.75 cloth; 50p. paperback. Pp. xii+144.

Dicks, D. R. [1970]: *Early Greek Astronomy to Aristotle.* London: Thames & Hudson. £2.50. Pp. 272.

Egidi, Rosaria [1971]: *Studi di logica e filosofia della scienza.* Rome: Biblioteca di Cultura, Bulzoni Editore. Lire 2,700. Pp. 255.

Emmet, Dorothy and MacIntyre, Alasdair (*eds.*) [1970]: *Sociological Theory and Philosophical Analysis.* New York: The Macmillan Co. £3.50. Pp. xxiv+232.

Fuller, Watson (*ed.*) [1971]: *The Social Impact of Modern Biology.* London: Routledge & Kegan Paul. £1.00 cloth; 50p. paperback. Pp. vii+256.

Goodstein, R. L. [1971]: *Development of Mathematical Logic.* London: Logos Press Ltd. £3.50. Pp. vii+150.

Graves, John C. [1971]: *The Conceptual Foundations of Contemporary Relativity Theory.* Cambridge, Mass. and London: MIT Press. £7.00, $15.00. Pp. ix+361.

Grene, Marjorie (*ed.*) [1971]: *Interpretations of Life and Mind: Essays around the Problem of Reduction.* London: Routledge & Kegan Paul. £2.25. Pp. xvi+152.

Hahn, Roger [1971]: *The Anatomy of a Scientific Institution: The Paris Academy of Sciences, 1666–1803.* Berkeley, Los Angeles, London: University of California Press Ltd. £5.70. Pp. xiv+433.

Hamlyn, D. W. [1971]: *The Theory of Knowledge,* Modern Introductions to Philosophy. London: Macmillan Press. £2.95 cloth; £1.50 paperback. Pp. x+308.

Handy, Rollo [1970]: *The Measurement of Values.* St. Louis, Missouri: Warren H. Green, Inc. $12.50. Pp. ix+221.

Hare, R. M. [1971]: *Essays on Philosophical Method: New Studies in Practical Philosophy.* London: Macmillan Press. £1.95. Pp. 135.

Hare, R. M. [1971]: *Practical Inferences: New Studies in Practical Philosophy.* London: Macmillan Press. £1.95. Pp. 120.

Haslip, Joan [1971]: *Imperial Adventurer.* London: Weidenfeld & Nicolson. £3.75. Pp. 544.

Heisenberg, Werner [1971]: *Physics and Beyond.* London: George Allen & Unwin Ltd. £3.00. Pp. viii+256.

Hirst, K. E. and Rhodes, F. [1971]: *Conceptual Models in Mathematics* (Mathematical Studies 5). London: George Allen & Unwin. £3.00. Pp. xviii+182.

Hooykaas, R. [1970]: *Catastrophism in geology.* Amsterdam, London: North-Holland Publishing Company. D.fl. 7.50. Pp. 50.

Hormann, Hans [1970]: *Psycholinguistics: An Introduction to Research and Theory.* Berlin, Heidelberg, New York: Springer-Verlag. Cloth DM 68, U.S. $18.70.

Hubbeling, H. G. [1971]: *Language, Logic and Criterion.* Amsterdam: Born N.V. $8.50. Pp. 128.

Jánossy, L. [1971]: *Theory of Relativity based on Physical Reality.* Budapest: Akadémia Kiadó. L5.00. Pp. 317.

Jardine, Nicholas and Sibson, Robin [1971]: *Mathematical Taxonomy.* London: John Wiley & Sons Ltd. £6.00. Pp. xviii+286.

Jorpes, J. Erik [1970]: *Jac. Berzelius His Life and Work* (translated from Swedish ms. by Barbara Steele). Berkeley and Los Angeles: Univ. of California Press. Pp. 156.

Kamenka, E. (*ed.*) [1970]: *A World in Revolution?* Canberra: Australian National University Press. $0.85. Pp. 100.

Klüver, Jürgen [1971]: *Operationalismus. Kritik und Geschichte einer Philosophie der exakten Wissenschaften.* Stuttgart: Friedrich Fromman Verlag (Reihe problemata frommann-holzboog). DM 19.80 paper; DM 32 cloth. Pp. 220.

Kuhns, Richard [1971]: *Structures of Experience: Essays on the Affinity between Philosophy and Literature.* New York; London: Basic Books. $7.95. Pp. 320.

Larson, Dewey B. [1971]: *Quasars and Pulsars.* Portland, Oregon: North Pacific Publishers. $6.50. Pp. x+180.

Lenk, H. (*ed.*) [1971]: *Neue Aspekte der Wissenschaftstheorie.* Braunschweig: Friedr. Vieweg & Sohn; Oxford & New York: Pergamon Press. Pp. 249.

Levensky, Mark (*ed.*) [1971]: *Human Factual Knowledge* (*Central Issues in Philosophy Series*). New Jersey: Prentice-Hall Inc. $6.95 cloth; $2.95 paperback. Pp. vi+218.

Levere, Trevor H. [1971]: *Affinity and Matter: Elements of Chemical Philosophy 1800–1865.* London: Oxford University Press. £4.50. Pp. xiii+230.

Linderholm, Carl E. [1971]: *Mathematics Made Difficult.* London: Wolfe Publishing Ltd. £2.75. Pp. 207.

Lorenz, Konrad [1971]: *Studies in Animal and Human Behaviour,* 2. London: Methuen & Co. Ltd. £4.25. Pp. xii+366.

McCormmach, Russell [1970]: *Historical Studies in the Physical Sciences,* 1. Philadelphia: University of Pennsylvania Press. $8.50. Pp. ix+314.

McCormmach, Russell [1970]: *Historical Studies in the Physical Sciences,* 2. Philadelphia: University of Pennsylvania Press. $8.50. Pp. ix+314.

MANSFIELD, J. [1971]: *The Pseudo-Hippocratic Tract Ch. 1–11 and Greek Philosophy.* Assen: Royal Vangorcum. dfl. 42.50 cloth. Pp. viii+271.

MARSHALL, Bernice (*ed.*) [1971]: *Experiences in Being*. Belmont, California: Brooks/Cole Publishing Co. (A Division of Wadsworth Publishing Co. Inc.) [English distributor: Prentice Hall]. £2.50. P. 317.

MERLEAU-PONTY, Jacques and MORANDO, Bruno [1971]: *Les Trois Etapes de la Cosmologie.* Paris: Editions Robert Laffont. Paperback. Pp. 316.

MICHALOS, Alex C. [1971]: *The Popper-Carnap Controversy*. The Hague: Martinus Nijhoff. D.fl. 22.50. Pp. 124.

MITTELSTRASS, Jürgen [1970]: *Neuzeit und Aufklärung*. Berlin, New York: De Gruyter. DM 68. Pp. xiii+651.

MORLEY, Lawrence [1971]: *The Progressive Anarchist.* Wakefield: Onex Publications. Pp. xiv+221.

MORRISON, Denton E. and HENKEL, Ramon E. (*eds.*) [1970]: *The Significance Test Controversy: A Reader; Methodological Perspectives.* Chicago: Aldine Publishing Company. Pp. xviii+333.

MUNDLE, C. W. K. [1971]: *Perception: Facts and Theories.* Oxford, London, New York: Oxford University Press. £1.25 hardback. 55p. OPUS paper edition. Pp. 192.

OLSON, R. [1971]: *Science as Metaphor: The Historical Role of Scientific Theories in Forming Western Culture.* Belmont, California: Wadsworth Publishing Co. Inc. [English distributor: Prentice Hall]. £3.00. Pp. ix+321.

OWEN, George E. [1971]: *The Universe of the Mind.* Baltimore and London: The Johns Hopkins Press Ltd. £7.15. Pp. xiv+349.

PITCHER, George [1971]: *A Theory of Perception.* Princeton: University Press; London: Oxford University Press. £4.00. Pp. 238.

PASTORE, Nicholas [1971]: *Selective History of Theories of Visual Perception 1650–1950.* New York, London, Toronto: Oxford University Press. £4.50. Pp. 454.

PEACOCKE, A. R. [1971]: *Science and the Christian Movement.* London: Oxford University Press. £4.00 cloth; £1.25 paperback. Pp. xiii+214.

RAO, A. P. [1971]: *Quine's Criterion of Ontological Commitment.* Simla: Indian Institute of Advanced Study. Pp. 28.

RAVETZ, J. R. [1971]: *Scientific Knowledge and its Social Problems.* Oxford: Clarendon Press. £5.00. Pp. x+449.

REDMAN, L. A. [1971]: *The Physics Teaching Handbook.* London: Longman Group Ltd. 75p. paperback. Pp. 232.

RENSCH, Bernhard [1971]: *Biophilosophy* (translated by Cecilia Sym) New York, London: Columbia University Press. £5.65. Pp. xii+128.

ROSENBLUETH, Arturo [1971]: *Mind and Brain: A Philosophy of Science.* Cambridge, Mass.: The MIT Press. $1.95 paperback. Pp. xii+128.

RUBINOFF, Lionel [1971]: *Collingwood and the Reform of Metaphysics.* Toronto: Toronto University Press; London: Oxford University Press. £6.00. Pp. xiv+413.

RUTLAND, W. R. [1971]: *The Becoming of God.* Oxford: Basil Blackwell. £1.75. Pp. xvii+122.

SCIAMA, D. W. [1971]: *Modern Cosmology.* Cambridge: Cambridge University Press. £3.60. Pp. 212.

SNYDER, D. Paul [1971]: *Modal Logic & its Applications.* London: D. Van Nostrand Co. Ltd.; New York, Cincinatti, Toronto, Melbourne: Van Nostrand Rheinhold Co. £6.25. Pp. xiv+335.

SPINOZA, Baruch de [undated]: *I Principi di Filosofia di Cartesio e l'Appendice* a cura di Bruno Widmar. Lecce: Edizioni Milella. L3,000 paperback. Pp. 213.

TAMÁS, György (*ed.*) [1971]: *Aufsätze Über Logik.* Budapest: Akadémiai Kiadó. L4,000. Pp. 288.

TEMPLE, G. [1971]: *The Structure of Lebesgue Integration Theory.* Oxford: Oxford University Press. £3.75. Pp. 184.

TISELIUS, Arne and NILSSON, Sam (*eds.*) [1971]: *The Place of Value in a World of Facts, Proceedings of the fourteenth Nobel Symposium.* New York, London, Sydney: John Wiley & Sons, Inc. £7.00. Pp. 496.

VAN HEIJENOORT, Jean [1970]: *Frege and Gödel: Two Fundamental Texts in Mathematical Logic.* Harvard: Harvard University Press; London: Oxford University Press. £2.90. Pp. 117.

VERMA, D. [1971]: *New Sex Values*. New Delhi: Star Publications; London: Independent Publishing Co. Inland Rs. £1.00 (Overseas £1.50). Pp. 116.

VON WRIGHT, George Henrik [1971]: *Explanation and Understanding* (International Library of Philosophy and Scientific Method). London: Routledge & Kegan Paul. £3.00. Pp. xvii+230.

WEINGARTNER, Paul [1971]: *Wissenschaftstheorie* 1. *Einführung in die Hauptprobleme.* Stuttgart: Friedrich Frommann Verlag (Reihe problemata frommann-holzboog). DM 19.80 paper; DM 32 cloth. Pp. 246.

WESTFALL, R. S. [1971]: *Force in Newton's Physics*. London: MacDonald; New York; American Elsevier. £10.00. Pp. xii+579.

WILMINGTON, Martin [1971]: *The Middle East Supply Centre.* New York: State University of New York Press. $10.00. Pp. xv+248.

WITTGENSTEIN, L. [McGUINESS, B. F., NYBERG, T. and VON WRIGHT, G. H. (*eds.*)] [1971]: *Prototractatus* An Early version of Tractatus Logico-Philosophicus by Ludwig Wittgenstein. London: Routledge and Kegan Paul. £7.50. Pp. 256.

Reason and Experience

The Representation of Natural Order in the Work of Carl von Linné

James L. Larson

Attempts to solve one of the oldest problems in Linnaean scholarship—the nature and extent of the subterranean connection between logical and natural forms implicit in Linné's system of nature.

208 pages illustrated £3.35

The Anatomy of a Scientific Institution

The Paris Academy of Sciences (1666-1803)

Roger Hahn

468 pages illustrated £5.70

University of California Press

Observation and Theory in Science

Contains: *Theory and Observation* by Ernest Nagel; *Science and Forms of Ignorance* by Sylvain Bromberger; and *Can We Ascertain the Falsity of a Scientific Hypothesis?* by Adolf Grünbaum.

134 pages £3.30

Concepts of Science

A Philosophical Analysis

Peter Achinstein

"... a book which no one can afford to miss who is at all interested in problems about the nature of theory construction in the sciences."—L. J. Cohen in *British Journal for the Philosophy of Science.*

280 pages Paperback £1.20

The Experimenters

A Study of the Accademia del Cimento

W. E. Knowles Middleton

The first extended study in English of the Accademia features a new translation of their *Saggi di naturale experienze* and assesses their contributions to the physical science of the seventeenth century.

416 pages Illustrations £5.70

The Johns Hopkins Press

2-4 Brook Street, London W1Y 3AA

Physics and Beyond

WERNER HEISENBERG

This is an intellectual autobiography in which the author recalls the development of atomic physics with which he was deeply involved from 1920 to 1965. It is in the form of dialogues and discussions with his contemporaries—scientists such as Planck, Einstein, Hahn, Fermi and others.
" . . . Heisenberg has scored impressively in the (cultural) gap-closing business by creating a work of art out of the memory of his life in science."
Elting Morison, New York Times Book Review

£3.00

Roots of Mankind

JOHN NAPIER

Dr. Napier approaches evolution from a functional point of view, by comparing the structure of the human body with that of fossils and of living primates. He postulates that the testing ground for all genetic innovation is the environment and that, *in this sense*, it has determined the direction of animal evolution.

£2.75: paper £1.75

Now in Paperback:

The Origins of Science

An Inquiry into the Foundations of Western Thought

ERNEST HUTTEN

£2.00

George Allen & Unwin

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY

Edited by Marvin Farber

DECEMBER, 1971, VOLUME XXXII, No. 2

CONTENTS

I ARTICLES
The Problem of Robot Consciousness, by DWIGHT VAN de VATE, JR.
Logical Impossibility, by SOLOMON E. LEVY
Factual Necessity and the Libertarian, by THEODORE GULESERIAN
Some Deontological Expressions, by CARL WELLMAN
Ontological Cryptophobia, by C. MASON MYERS
Sum is a Logical Consequence of *Cogito*, by RONALD SUTER

II DISCUSSIONS

III REVIEWS

IV NOTES AND NEWS

V RECENT PUBLICATIONS

Subscription prices $7.50 per year for individuals.

$9.00 per year for institutions.

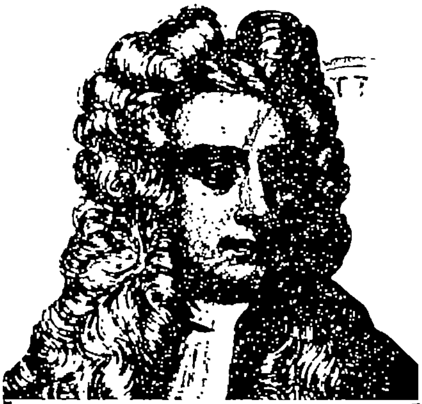
Separate issues, $2.00 and $2.50 respectively.

Postage and handling for back issues only.

Papers submitted for publication will not be returned unless accompanied by a self-addressed stamped envelope, or equivalent postage.

A circular listing the main contents of the Journal since 1940 will be sent upon request.

Address all communications to
Philosophy and
Phenomenological Research
State University of New York at Buffalo
Buffalo, New York 14226



Introduction to Newton's Principia

I. BERNARD COHEN

An introduction to the new edition of Newton's *Principia* which also describes how Newton came to write it, the stages of preparing the manuscript and seeing it through the press, the critical reception it received and the successive revisions Newton made for the second and third editions.
Published in the U.S.A. and Canada by Harvard University Press

£13.00 net

The Matter of Chance

DAVID HUGH MELLOR

The author develops a philosophical theory of chance (i.e. statistical probability) that makes sense for the first time of the four chief assumptions which are made in professional and everyday uses of the term; namely that it is objective, empirical, not relational and is applicable to the single case. Dr Mellor's thesis considerably refines and extends the relevant work of Braithwaite, Popper and Hacking, and relates it closely to personalist theories of probability.

£4.00 net

CAMBRIDGE
UNIVERSITY PRESS

INQUIRY An Interdisciplinary Journal of Philosophy and the Social Sciences
Edited by Arne Naess
and Alastair Hannay

Number 4 *Winter 1971* *Volume 14*

Toward a Theory of Moral Debt: (I) The Idea of Moral Debt in the Common Understanding	Morris B. Storer
Is 'Human Action' a Category?	Arthur B. Cody
Marx and the Origin of Dialectical Materialism	Stanley Moore
The Dialectics of Metaphor	David Bloor
Covering Laws in Historical Practice	Sten Sparre Nilson
Review Discussion: Glover on Responsibility	T. L. S. Sprigge

Published quarterly by Universitetsforlaget, P.O. Box 307
Blindern, Oslo 3, Norway

Annual Subscription: $7.50, £2·75 including postage

MIND

A Quarterly Review of Psychology and Philosophy
edited by Professor **DAVID HAMLYN**

JANUARY 1972

I	Scepticism and Private Language	E. R. WINKLER
II	Intentions, Beliefs, and Imperative Logic	R. F. STALLEY
III	Pattern Recognition and Scientific Progress	J. J. SPARKES
IV	The Foundations of Logic	B. H. SLATER
V	Professor Grice's Theory of Meaning	A. F. MACKAY
VI	The Truth or Falsity of Value Judgements	DOROTHY MITCHELL
VII	The Corrigibility and the Generality of Language	E. PIVČEVIĆ
VIII	Discussions:	
	Professor Armstrong on 'Non-Physical Sensory Items'	H. M. ROBINSON
	Standards and Criteria: Can God be the Standard of the Good?	T. C. MAYBERRY
	Confirmationism	D. G. COLLINGRIDGE
	A Defence of the No-Ownership Theory	D. S. CLARKE
	Levels of Rules and Hart's Concept of Law	G. GERBER
	Sensations and Processes—A Reply to Munsat	U. T. PLACE
	Visualizing	M. DEITSCH
	Smart's Identity Theory, Translation, and Incorrigibility	S. J. NOREN
	Fragmenting Subjects	C. G. PRADO
	A Note about Presupposition	F. H. DONNELL
	Frege, Geach, and 'The Concept *Horse*'	W. GUSTASON
	Quine on Dispositions and Subvisible Structures	J. MOLINE
	Foreknowledge and Predestination	L. BECKER
	A Note on Mr. Cooper's Reconstruction of Mill's "Proof"	N. GRIFFIN
IX	Critical Notice:	
	Jan Lukasiewicz: Selected Works. Edited by L. Borkowski.	W. KNEALE
X	New Books	
XI	Notes	

Annual Subscription £1.25, post free; single copy 40p, plus postage

Published for MIND ASSOCIATION by
BASIL BLACKWELL 49 BROAD STREET OXFORD

Noûs

Nihil philosophici a nobis alienum putamus

A Quarterly Journal of Philosophy

edited by Hector-Neri Castañeda

Philosophy Department, Indiana University

Bloomington, Indiana 47401

NOÛS is committed to the publication of first-rate essays and short discussions on philosophical problems regardless of the author's philosophical school or point of view. It includes critical studies of recently published books rather than reviews and critical summary surveys of studies as well as symposia dealing with specific philosophical problems.

A quarterly, NOÛS is published in March, May, September, and November each year. For *individuals* the subscription rates are: one year, $8.00; two years, $14.50; three years, $20.00. For *bona fide students*: $5.00 per year. For *institutions*: $12.00 per year. Add 15 cents per year to Canadian subscriptions and 35 cents per year to all other foreign subscriptions.

CONTENTS

Volume V, Number 4, November 1971

DONALD DAVIDSON—Eternal vs. Ephemeral Events

R. J. HAACK—On Davidson's Paratactic Theory of Oblique Contexts

LESLIE H. THARP—Truth, Quantification, and Abstract Objects

JOHN HOWARD SOBEL—Value, Alternatives, and Utilitarianism

LES HOLBOROW—The Commitment Fallacy

KEITH HALBASCH—A Critical Examination of Russell's View of Facts

RICHARD E. AQUILA—Internationality and Possible Facts

DOUGLAS LACKEY—Massey on Tense and Special Relativity

Publications Received

Order from
The Secretary, NOÛS
Department of Philosophy
702 Ballantine Hall
Indiana University
Bloomington, Indiana 47401

Volume 23 Number 2 May 197

The British Journal for the Philosophy of Science

CAMBRIDGE UNIVERSITY PRESS

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editor: IMRE LAKATOS

Assistant Editors: COLIN HOWSON, JOHN WORRALL AND ELIE ZAHAR

Subscriptions

The *Journal* is published quarterly. The subscription price is £4·50 net in U.K. (U.S.A. $14.00) per volume post free; single issues £1·50 net in U.K. (U.S.A. $4.75) plus postage. The subscription is £3·00 (U.S.A. $9.50) for members of the British Society for the Philosophy of Science, who should order direct from the publisher. Orders may be sent to any bookseller or direct to Cambridge University Press, Bentley House, 200 Euston Road, London NW1 2DB. Orders from U.S.A. and Canada should be sent to Cambridge University Press American Branch, 32 East 57th Street, New York, N.Y. 10022, U.S.A. Second class postage paid at New York, N.Y.

Notes to Contributors

Articles, correspondence and books for review should be sent to the Editor, Professor Imre Lakatos, at the Department of Philosophy, London School of Economics, Houghton Street, Aldwych, London WC2A 2AE. Articles should not normally exceed 5,000 words, and should be accompanied by abstracts of about 100 words.

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Footnotes, which should be as few as possible, should be typed separately, double-spaced at the end of the text. They should be referred to by consecutive numbering throughout the text, in which they will be inserted in page proof. Bibliographical references should be listed alphabetically at the end, and reference made to this list in the text rather than in footnotes. The style of references should follow current *Journal* practice.

Long quotations are printed in small type without quotes, and should be so marked in the typescript. Single quotes should be used to form the names of words and sentences, and double quotes for all other purposes. Diagrams, tables, and illustrations should be on separate sheets, with their desired position in the text clearly indicated. Symbolic formulae appearing as separate lines should be clearly distinguished in the text, and foreign letters (Greek, German, etc.) indicated in the margin. Material to be printed in italics, especially English letters used as symbols ('A', 'x', etc.), should be underlined. Numerals should not be italicised. On matters of style and layout, authors should consult CHAUNDY, T.W. *et al.* [1964]: *The Printing of Mathematics*, London: Oxford University Press.

Offprints of articles may be ordered when corrected galley proofs are returned. Fifty are supplied free ; the price of extra copies is given on the order form.

© Cambridge University Press, 1972

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the London or New York office of Cambridge University Press.

ISI Tear Services, 325 Chestnut Street, Philadelphia, Pennsylvania 19106, U.S.A. is authorized to supply single copies of separate articles for private use only.

Advertising

Particulars of advertising in the *Journal* may be obtained from Cambridge University Press, Bentley House, 200 Euston Road, London, NW1 2DB; or in U.S.A. and Canada, Cambridge University Press American Branch, 32 East 57th Street, New York, N.Y. 10022, U.S.A.

Back Volumes

Volumes 1-16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17-21 may be obtained from Cambridge University Press.

Brit. J. Phil. Sci. 23 (1972), 89–109 *Printed in Great Britain*

Two Deviant Logics for Quantum Theory: Bohr and Reichenbach*

by MICHAEL R. GARDNER

1 *The Paradoxes.*
2 *The Copenhagen Interpretation.*
3 *Bohr Revised: Three-Valued Logic.*
4 *Measurement: The Projection Postulate.*
5 *Measurement: The Uncertainty Principle.*
6 *New Paradoxes.*
7 *Conclusion: Bohr Reconsidered.*

1 THE PARADOXES

Foundational studies in quantum theory have consisted in large part of attempts to avoid certain paradoxes. This paper concerns Bohr's and Reichenbach's attempts to avoid some of these seeming absurdities by altering quantum theory's logic rather than its specifically physical postulates. (A different logical change to the same end—a change involving the distributive laws—is discussed elsewhere (Gardner [1971]), along with solutions to the paradoxes which involve no changes in logic. The paradoxes themselves are described in somewhat more detail there.)

The initial group of paradoxes we shall discuss consists of the two-slit paradox, the tunneling paradox, and the orbital electron paradox. Later we shall discuss the Schrödinger cat paradox and a modified version of the Einstein-Podolsky-Rosen paradox and show how they defeat Bohr's and Reichenbach's solutions to the first three.

The two slit paradox concerns an experimental set-up in which photons from a light source are allowed to pass through either of two slits in a diaphragm, and then strike a photographic plate behind it. Let $P(A_i, R)$ be the probability that a photon which passes through slit i ($i = 1, 2$) strikes a small region R on the plate. Let $P(A_i)$ be the probability that a photon emitted by the source passes through i, and suppose the emission pattern is symmetric so that $P(A_1) = P(A_2) = P(A_1 \vee A_2)/2$.

---

* Part of a doctoral dissertation, Harvard University, 1971. I wish to express thanks to Jeffrey Bub, Hilary Putnam, Abner Shimony, Lawrence Sklar, and John Stachel for helpful discussions and criticisms of earlier drafts. My general point of view owes much to Ballentine [1970].

Received 1 *May* 1971

$$\begin{aligned} \text{Then}\quad P(A_1 \vee A_2, R) &= P((A_1 \vee A_2) . R)/P(A_1 \vee A_2) \\ &= P(A_1 . R \vee A_2 . R)/P(A_1 \vee A_2) \\ &= P(A_1 . R)/P(A_1 \vee A_2)+P(A_2 . R)/P(A_1 \vee A_2) \\ &= P(A_1 . R)/2P(A_1)+P(A_2 . R)/2P(A_2) \\ P(A_1 \vee A_2, R) &= \tfrac{1}{2}P(A_1, R)+\tfrac{1}{2}P(A_2, R) \end{aligned}$$

We can measure $P(A_1 \vee A_2, R)$ by counting the number of photons striking R in a long time-interval and dividing by the number striking the entire screen—*i.e.* the number passing through either 1 or 2. Since the value of $P(A_1, R)$ should presumably be unaffected by closing 2, we can measure it by counting the arrivals during another interval during which only 1 is open. $P(A_2, R)$ is measured with only 2 open. The paradox is that the values thus obtained do not satisfy the equation derived above (following Putnam [1970]; cf. Heisenberg [1958]).

The second paradox relates to the quantum-mechanical phenomenon of tunneling, or barrier-penetration. Suppose that a particle with total energy E less than V_0, the value classically required to escape from a potential well, is initially inside the well. According to quantum mechanics, the probability is positive that the particle will later be found outside the well, even though to arrive there it must pass through a region in which its potential energy exceeds its total energy, which is impossible. The escape of alpha particles from uranium nuclei provides experimental confirmation of this result. The energy E of the escaped particles may be measured in the region $x = \infty$ (where $V(x) = 0$ and $E =$ the kinetic energy) and the paradoxical relation $E < V_0$ thereby verified (modified version, from Reichenbach [1944], p. 165).

The orbital-electron paradox is due to Heisenberg ([1930], p. 33). If a hydrogen atom is in an eigenstate of the Hamiltonian and the eigenvalue is E, the electron presumably cannot move out beyond the point at which the monotonically increasing potential energy V is equal to E. But according to quantum theory, the probability is positive that the electron will be found at any arbitrarily large distance from the nucleus. In any sufficiently large sample of atoms with energy E, there will therefore be some in which the potential energy exceeds the total energy, just as in the second paradox.

2 THE COPENHAGEN INTERPRETATION

The solution to such paradoxes which is almost universally accepted by physicists is called 'the Copenhagen interpretation of quantum theory', and is associated most closely with the name of Niels Bohr. He reasoned that 'the unavoidable interaction between the objects and the measuring instruments sets an absolute limit to the possibility of speaking of a behavior of atomic objects which is independent of the means of observation' (Bohr [1958], p. 25). Accordingly, it is impossible to make 'any sharp separation between the behavior of atomic objects and the interaction with the measuring instruments' (*ibid.*, p. 40). (Heisenberg stated the view in more precise language: 'the concept of the probability function does not allow a description of what happens between observations. Any attempt to find such a description would lead to contradictions' (Heisenberg [1958], p. 52).) Though section 7, below, will suggest that it may be uncharitable to do so, for the present we shall interpret Bohr as proposing that if at a time t, no measurement is occurring of an observable belonging to an atomic object, then any sentence concerning a value of that quantity at that time is meaningless. We will view this as a proposal to adopt a rather peculiar definition of a well-formed formula, and hence as a proposal for a deviant logic for quantum theory.

The exclusion from quantum theory of certain classical formulas as ill-formed enables one to resolve the three paradoxes without much additional difficulty. In deriving the two-slit paradox, one must assume that the sentences A_i, 'The photon passed through slit i,' ($i = 1, 2$), are meaningful. However, since the experiment does not provide any means to measure the position of the photon as it passes the diaphragm, these sentences are meaningless on Bohr's view. (Bohr also argued that if we did determine which slit each photon passed through, the interaction of the measuring apparatus with the photons would destroy the paradoxical distribution (*ibid.*, p. 46).)

The barrier-penetration paradox may be dealt with in the same spirit. In a case where no measurement of E occurs until the particle has escaped into the region where $V = 0$, sentences concerning the values of E and $V(x)$ in the classically forbidden region are meaningless. The paradoxical sentence '$E < V(x)$,' which seemed to follow from the prediction that tunneling will occur, is actually ill-formed. If, on the other hand, the particle undergoes measurement in the classically forbidden region, a follower of Bohr must invoke the Heisenberg uncertainty principle. According to this principle on its usual interpretation, it is impossible to perform an (exact) measurement of simultaneous values of x and p_x. Therefore, any sentence ascribing a value to $E = (p_x^2/2m) + V(x)$ at a position x is always meaningless, since a measurement of x precludes a measurement of p_x and, *a fortiori*, one of $(p_x^2/2m) + V(x)$. The solution of the orbital-electron paradox is essentially the same, since the argument sustaining it also requires reference to the energy at a particular position.

A careful reading of Bohr discloses a contrasting strand in his thinking, in which he seems to be advocating an interpretation distinct from that of Heisenberg. On Heisenberg's view, it is meaningless to speak of the properties of a system not undergoing measurement. But in his famous reply to Einstein, Podolsky and Rosen, Bohr conceded that if the positions and momenta of two particles were correlated in a certain way because of a previous interaction, the position or momentum of one might be determined by a measurement on the second (Bohr [1935], p. 699). Accordingly, there would be a meaningful and indeed true ascription of a position or momentum to the first particle, even though it was not itself interacting with a measuring device. Bohr's only disagreement with Einstein *et. al.* concerned their contention that in such a case, the first particle would possess both a position and a momentum, because one could determine either magnitude (but not both). Bohr felt that since a determination of one of these magnitudes would render impossible a prediction of the result of determining the other, a particle could possess at most one of the conjugate magnitudes at a given time (Bohr [1935], p. 700; cf. his [1958], p. 87). This argument was evidently based on the assumption that if the result of measuring a quantity cannot be predicted with certainty—that is, if the system is not in an eigenstate of the corresponding operator—then no statement about its value is true or even meaningful. As he put it, the 'conditions which define the possible types of predictions regarding the future behaviour of the system . . . constitute an inherent element of the description of any phenomenon to which the term "physical reality" can be properly attached'. Conversely, an eigenstate is a sufficient condition of truth on this view; for Bohr implies elsewhere that an electron not undergoing measurement nonetheless has a mass, 'since measurements of its inertial mass always give the same result' (Bohr [1958], p. 98). That meaninglessness rather than falsity is involved is suggested by his assertion that the uncertainty principle, which excludes simultaneous eigenstates of position and momentum, also imposes a 'limitation of the well-defined application of space-time concepts and dynamical conservation laws' (Bohr [1963], p. 5). That is, the absence of an eigenstate of a quantity A entails the inapplicability of the concept of A, the meaninglessness of statements of the form '$A = r$,' where r is a real number.

It will presumably be obvious how to resolve the paradoxes using the second type of Copenhagen view. (The resolution anyway mirrors Reichenbach's solutions in the next section. One need only substitute 'meaningless' for 'indeterminate.')

3 BOHR REVISED: THREE-VALUED LOGIC

Many objections might be raised to the Copenhagen versions of quantum theory. For the present, we shall confine ourselves to one. This will suggest a revision, which will then be criticised on other grounds, some of which will be applicable to the unrevised versions as well.

The objection in question is that either theory embodies a certain formal awkwardness—specifically, that well-formedness is not a purely syntactic property. A sentence of the form '$x = a$ at t' will be meaningful for some values of the time t but not for others, depending upon whether any measurement of x occurs at t, or upon the system's state at t (cf. Reichenbach [1944], p. 143). This requirement has the consequence that if t is in the distant past, we may be unable to decide whether '$x = a$ at t' is meaningful unless we carry out an extensive historical investigation to determine whether x was measured at t, or what the system's state was then. If t is in the distant future, the decision as to meaningfulness may have to be postponed until time t arrives. This awkwardness makes it impossible to formulate Bohr's theory as anything remotely resembling a formal system with a recursive class of well-formed formulas.

For reasons like this, Reichenbach proposed to admit sentences about unmeasured quantities as meaningful, but to ascribe to them a third truth-value, which he called 'indeterminacy'. Replacing 'meaningless' in Bohr's second theory with 'indeterminate', Reichenbach adopted the following 'conventions' or 'definitions':

Definition 4. The result of a measurement represents the value of the measured entity immediately after the measurement.

Definition 5′. In a physical state not preceded by a measurement of an entity u, any statement about a value of the entity u is indeterminate. (implicit in his [1944], pp. 140–5)

Reichenbach's theory of measurement suggests a re-statement in terms of wave functions of his conditions for the ascription of the three truth-values. He remarks that 'the interval "immediately after" of Definition 4 is determined by the time dependence of the Ψ-function; in a stationary state this interval may be chosen as long as we wish' (*ibid.*). He also defines a measurement of an entity u as 'a physical operation relative to which the Ψ-function of the physical system is [at least approximately[1]] one of the eigenfunctions of u' (*ibid.*, p. 97). If the measurement yields the value u_i, then the Ψ-function after the measurement must be one corresponding to the eigenvalue u_i (*i.e.* $u\Psi = u_i\Psi$). Relative to this theory of measurement, Definitions 4 and 5′ are equivalent to the following conditions for ascription of the three truth-values[2]:

(1) '$u = u_i$' is true $\equiv u\Psi = u_i\Psi$

---

[1] Henceforth this qualification will be ignored. We shall also pretend that continuous observables have eigenfunctions.

[2] 'u' refers elliptically to the value of the observable (operator) u in a system with wave-function Ψ. This interpretation of Reichenbach is due to van Fraassen.

(2) '$u = u_i$' is false $\equiv u\Psi = u_j\Psi$, for some $u_j \neq u_i$

(3) '$u = u_i$' is indeterminate $\equiv \Psi$ is not an eigenstate of u.

The admission of a third truth-value requires new definitions of the propositional connectives. The only one of interest to us is disjunction, which is defined as true when at least one disjunct is true, false when both are false, and indeterminate otherwise. This alone allows us to dissolve the two-slit paradox. The argument supporting the paradox requires the premise that any photon striking the plate passed through one slit or the other: $A_1 \vee A_2$. However, since none of the photons undergoes a position measurement (*i.e.* none is in an eigenstate of position) at the diaphragm, A_1 and A_2 are indeterminate, and so therefore is '$A_1 \vee A_2$' (*ibid.*, pp. 160 ff.). But if a statement is indeterminate, no one can be in a position to assert it, since no one can know that it is true. It follows that no one can prove anything, paradoxical or otherwise, by means of an argument in which such a statement figures as a premise.

An immediate consequence of (1)–(3) and the uncertainty principle enables us to resolve the second and third paradoxes. No wave function can be an eigenfunction of both x and p_x. Therefore, if '$x = x_i$' is true, then any statement concerning a value of p_x is indeterminate. The derivation of the tunneling paradox requires us to assert that $E = (p_x^2/2m) + V(x)$ has the same value at every position x. But this is a statement about simultaneous values of p_x and x, and is therefore never assertable (cf. *ibid.*, p. 166). The orbital electron paradox is dissolved by noting that if the wave-function is an eigenfunction of energy, then it is not an eigenfunction of position. Consequently, the premises concerning the electron's position are indeterminate and hence not assertable.

4 MEASUREMENT: THE PROJECTION POSTULATE

Reichenbach's version of quantum mechanics is based upon the projection postulate: the assertion that if a measurement of an observable u in a system S yields the value u_i, then the state of S after the measurement is an eigenstate of u with eigenvalue u_i. If Ψ before the measurement is $\Sigma a_j \Psi_j$, then the process of measurement 'reduces' it to just Ψ_i, where $u\Psi_i = u_i\Psi_i$. One difficulty with this view is that the postulated reduction of the wave packet violates the Schrödinger equation, according to which no discontinuous and unpredictable transformation of the wave function should occur. We therefore seem forced to conclude that a measurement is a peculiar sort of interaction which is not governed by the laws governing all other processes. But since measurements are just interactions among physical systems, it is very strange that there should be this unexplained difference between them and other interactions.[1] To provide such an explanation, some writers have suggested that the Schrödinger equation is not correct, and that it should be supplemented by certain additional terms which can be shown to accomplish reduction of the wave packet during measurement (cf. Bohm and Bub [1966]; Wigner [1963]).

The difficulties to which the projection postulate leads suggest that it may be worthwhile to re-examine critically the reasons for which it has been adopted. In an early work ([1930], p. 36) Heisenberg argues:

> The mathematical representation of the physical process changes discontinuously with each new measurement; the observation singles out of a large number of possibilities . . . the one which has happened. . . . As our knowledge of the system does change discontinuously at each observation its mathematical representation must also change discontinuously.

Von Neumann in his [1932], p. 214, argues similarly:

> If the system is initially found in a state in which the values of R cannot be predicted with certainty, then this state is transformed by a measurement M of R . . . into another state: namely, into one in which the value of R is uniquely determined [*i.e.* can be predicted with certainty].

This latter state, of course, must be an eigenfunction of R. Noting that this type of transformation violates the Schrödinger equation, von Neumann remarks that the violation 'remains unexplained for the present,' but asserts that he will 'attempt to bridge this chasm later' (*ibid.*, pp. 217 f.).

It is evident from Heisenberg's and von Neumann's arguments for the projection postulate that they base it upon a subjective interpretation of probability and therefore of the wave function. Ψ represents our knowledge and ignorance of the system, what we can and cannot predict with certainty. But this conception of quantum-mechanical probability is evidently untenable. If the 'we' is intended to refer to each of us individually, then, since we each possess a different body of knowledge and of predictive capabilities, does a system have a different wave function relative to each person? If I perform a measurement and inform a friend of the result the next day, does the wave function relative to him change a day later than the one relative to me? On the other hand, if 'our knowledge' is intended—as is perhaps more likely—to refer to that of the scientific community as a whole, then Heisenberg and von Neumann would still face the difficulty that when one of the community's members performs a measurement, the community's total knowledge and hence the wave function which allegedly represents it undergo unpredictable and discontinuous changes, in apparent violation of quantum theory.

---

[1] As Einstein remarked, 'measurements can occur only as special instances, *viz.*, parts of physical description, to which I cannot ascribe any exceptional position above the rest' (Einstein [1949], p. 674).

Popper has suggested that we can avoid difficulties of these sorts by adopting a certain objective interpretation of probability—one making no reference to any person's or group's knowledge. According to the relative frequency interpretation, the probability that x is an A is the limiting frequency of A's is some infinite sequence B of which x is a member. This probability depends upon the choice of B—the 'reference class'—and should therefore be labelled 'probability$_B$'. Popper, in his 'propensity interpretation of probability', has suggested that at least in quantum mechanics, the only appropriate choices for B are sequences which would be produced by an infinite number of realisations of certain physical 'generating conditions'—*i.e.*, by repetition of an experiment (cf. Popper [1967], pp. 32 ff.). The experiment might be, for example, sending an electron through a single slit in a diaphragm. The probability$_B$ that such an electron will appear in a region R on a screen would then be the limit (if it exists) which R_n/n would approach if the experiment were repeated indefinitely many times, where R_n is the number of electrons which would arrive at R if the first n members of B passed through the slit.

If this analysis of quantum-mechanical probability is correct, it resolves our difficulties concerning the projection postulate. If an electron has passed through a slit but has not yet arrived at the screen, we do not know whether it will arrive at R. After it arrives there and we look at the screen, we do know. But this knowledge of ours does not affect the probability that an electron which passes through the slit arrives at R; for this probability is simply a certain limiting ratio in an infinite sequence, and this ratio obviously does not change when we find out what one of the members of the sequence is. Therefore, on this view, no reduction of the wave packet occurs; *i.e.* the projection postulate is false. (This is the point of view taken by Popper in his [1935], ch. ix.)

Popper has recently suggested (see his [1967], p. 34) that it is possible to resurrect the projection postulate by interpreting it as corresponding to a change of reference class. When we learn that an electron arrives at a point P on the screen, we may decide to make use of this new information in order to predict its future course. If so, we will be interested in the probability that a particle which arrives at P will arrive at a farther point Q, rather than the probability that a particle which passes through the slit will arrive at Q. Accordingly, we will replace the electron's original wave function with a new one calculated by using this additional information. Initially, this will be an eigenfunction of position, since if a second position measurement is taken quickly enough, it will yield a position arbitrarily close to P. This new wave function will change in accordance with the Schrödinger equation, just as the old one did before the measurement.

At least in the case of measurements which are repeatable in the above sense, the original wave function $\Psi = \Sigma a_i\Psi_i$ may be replaced by Ψ_i, where Ψ_i is the eigenfunction (of the measured observable) whose eigenvalue is the result of the measurement. However, since this change corresponds to a shift in reference class, and since the Schrödinger equation determines the temporal development of a wave function chosen by specifying a *single* reference class, it is a mistake to think that this 'projection' violates the Schrödinger equation, and to wonder as Shimony does in his [1963] how the human mind could have the strange power to cause these violations. *A fortiori* it is a mistake to follow Bohm and Bub (cf. their [1966]) and Wigner (cf. his [1963]) in thinking that the phenomenon of projection requires modification of the Schrödinger equation. Projection is something *we* do, at our option, and at a time of our own choosing. It is therefore not to be expected that the process will have a counterpart in the basic time-dependent equations governing the motions of microscopic particles.

It would also be a mistake to conclude, however, that so long as it is understood as corresponding to a change in reference class, the projection postulate is suitable as a general description of measurement. First of all, there are many measurements which alter the value of the observable measured. An immediate repetition would not yield the same result, and the measurement therefore is not a preparation of an eigenstate corresponding to the eigenvalue obtained. For example, a standard device for measuring the energy and direction of scattered particles is the solid-state detector. It functions by removing nearly all of the energy from the particle, and emitting a voltage pulse whose height is proportional to the energy removed. At the end of the measurement, the particle has neither the direction nor the energy it had when it entered the aperture; for the direction has been changed randomly by diffraction at the aperture and by collisions inside the detector, and the energy has been reduced nearly to zero. Similarly, as Popper has pointed out ([1967], p. 27), if we use a magnetic spectrograph to determine the velocity of an alpha particle, the velocity is altered randomly when the particle strikes the photographic plate. But the position at which it strikes nonetheless allows us to calculate its velocity during the entire time it was in the magnetic field.

The projection postulate is not suitable, then, as a *general* description of measurement. The most one would seem entitled to claim is that if the measurement is of the repeatable kind (*e.g.* of position or spin), then it prepares an eigenstate with the eigenvalue measured. However, a device M for performing repeatable measurements of an observable A may be constructed in such a way that it emits all systems which enter it regardless of the value of A it determines them to possess. Thus, even though repetition of a measurement of A upon an individual system S might yield the same value which M obtained for S, M will not in general constitute a preparation-device for any eigenstate of A, since not all the systems it emits will have the same value of A. Theorists who wish to specify the post-measurement state of a system relative to its emission from M, but not relative to M's having found a particular value of A, sometimes claim that a measurement of A prepares, not an eigenstate of A, but a ***mixture*** of such states.[1] In particular, one might assert that if the ensemble entering M has the wave-function $\Psi = \Sigma\langle\alpha_k|\Psi\rangle\alpha_k$ and hence density operator $\rho = P_\psi$, where the α_k's are the eigenvectors of A, then the ensemble emitted by M has the density operator

$$\hat{\rho} = \sum_k |\langle\alpha_k | \Psi\rangle|^2 P\alpha_k.$$

Park in his [1968*b*] has shown that this transformation is not a general characterisation of measurement. (That it is not, of course, is suggested by its relation to the assumption of repeatability.) However, it may be true of at least some measurements. The question then arises whether in special cases of measurements it is possible to derive the transformation to a mixture from the usual axioms of quantum theory, so that we do not have to postulate that measurements obey laws different from those obeyed by other interactions.

To answer this question (following Park [1968*a*, *b*]), let H_1 be the Hilbert space associated with the systems S undergoing measurement and H_2 with the measuring device M. The tensor product space $H = H_1 \otimes H_2$ is then associated with the composite system. Let $\{\gamma_k\}$ be a set of eigenstates of some observable C in M. $\{\alpha_k\}$ spans H_1, $\{\gamma_k\}$ spans H_2, and $\{\alpha_k \otimes \gamma_l\}$ spans H. If S is initially in Ψ and M is initially in χ, then $S+M$ is initially in $\Psi \otimes \chi$. We now make the specialising assumptions which permit the derivation of the transformation to a mixture: (1) if S is in an eigenstate α_k of A before the measurement, its state is unchanged by the measurement; and (2) the post-measurement state γ_k of M is correlated with the state α_k of S. If T is the temporal evolution operator for this interaction, these assumptions amount to:

$$T(\alpha_k \otimes \chi) = \alpha_k \otimes \gamma_k.$$

It follows that

$$T(\Psi \otimes \chi) = T(\sum_k\langle\alpha_k \mid \Psi\rangle\alpha_k \otimes \chi) = \sum_k\langle\alpha_k \mid \Psi\rangle\alpha_k \otimes \gamma_k \equiv \phi.$$

---

[1] von Neumann sometimes takes this position, but seems not to distinguish it from the projection postulate (cf. his [1932], p. 418). See also, *below*, p. 99, footnote 1.

Let $\hat{\rho}$ be the post-measurement density operator for $S+M$. Evidently, $\hat{\rho} = P_{\phi}$. The post-measurement density operator $\hat{\rho}_s$ for S alone may be obtained by taking a 'partial trace':

$$\begin{aligned}
\hat{\rho}_s &= Tr_M\hat{\rho} \\
&= \sum_l \langle\gamma_l \mid \hat{\rho} \mid \gamma_l\rangle \\
&= \sum_l \langle\gamma_l \mid \sum_k \langle\alpha_k \mid \Psi\rangle\alpha_k \otimes \gamma_k\rangle\langle \sum_m \langle\alpha_m \mid \Psi\rangle\alpha_m \otimes \gamma_m \mid \gamma_l\rangle \\
&= \sum_{lkm} \delta_{lk}\delta_{ml}\langle\alpha_k \mid \Psi\rangle\langle\alpha_m \mid \Psi\rangle^* \mid \alpha_k\rangle\langle\alpha_m \mid \\
&= \sum_k \langle\alpha_k \mid \Psi\rangle\langle\alpha_k \mid \Psi\rangle^* \mid \alpha_k\rangle\langle\alpha_k \mid \\
\hat{\rho}_s &= \sum_k \mid \langle\alpha_k \mid \Psi\rangle \mid^2 P_{\alpha_k}.
\end{aligned}$$

Similarly,

$$\hat{\rho}_M = Tr_s\,\hat{\rho} = \Sigma \mid \langle\alpha_k \mid \Psi\rangle \mid^2 P_{\gamma_k}.$$

We see, then, that the transformation of the state of the measured system to a mixture of eigenstates of the measured observable may be derived in special cases from the fundamental axioms of quantum theory.[1] This result will be important in section **6**.

Three paragraphs back, we entertained the worry that the transformation of a state produced by a measurement violates the Schrödinger equation and that measurements are therefore unexplainably different from other interactions. We have alleviated the second part of this worry by showing that it follows from quantum theory—specifically, from the linearity of the temporal evolution operator—that interactions of a certain kind will transform a pure state into a mixture of eigenstates of the measured observable. That interactions of that kind—*i.e.* that some measurements—should produce this change is therefore not unexplainable. However, the fact that this changes violates the Schrödinger equation may seem worth worrying about in its own right. Presumably, we ought to be able to include the interaction between S and M in the Hamiltonian H of S. If the density operator $\rho_s(0)$ of S at time 0 is $P_\psi(0)$, then the Schrödinger equation implies that

$$\Psi(t) = e^{-itH}\psi(0)$$

---

[1] The treatment above actually originates with von Neumann (cf. his [1932], pp. 422 ff.). Our disagreement with him, then, relates not to it but to (1) his failure to distinguish a measurement's producing eigenstates from its producing a mixture; (2) his failure to recognise that the former corresponds to a change of reference class and therefore requires no further explanation; and (3) his failure to note that the latter occurs only in special cases. (1) would not be an objection if the ignorance interpretation of mixtures were correct. But see section **6**.

and

$$\rho(t) = P_{\Psi(t)}.$$

Regardless of the precise nature of the interaction between S and M, then, the Schrödinger equation requires that the initially pure state of S remain pure. Quantum theory would therefore seem inconsistent with the occurrence of measurements of the type considered above.

To resolve this antinomy we must examine the derivation of the Schrödinger equation in order to determine the limits of its applicability. As Mackey shows (see his [1963], pp. 81 ff.), if we assume that the state of a system at time $t_2 > t_1$ is uniquely determined by $t_2 - t_1$ and the state at time t_1, then it is possible to show, given certain linearity and continuity assumptions, that there is an operator H such that for all pure states $\Psi(0)$,

$$\Psi(t) = e^{-itH}\Psi(0).$$

It follows that

$$\frac{d\Psi}{dt} = -iH\Psi.$$

The operator H, of course, is called the *Hamiltonian.*

This derivation requires the assumption that the way in which the state of a system changes between two times does not depend on the times themselves but only on the difference between them. This is a reasonable assumption for an isolated system (or for one influenced only by static external forces), and applies in particular to $S+M$ if it is isolated. As we saw above, the interaction therefore results in no loss of purity for the state of the composite system. However, Mackey's assumption is clearly inapplicable to S alone. The temporal evolution of S's state during a given time-interval obviously depends upon whether S interacts with M during that interval, and is therefore not determined solely by the length of the interval. Accordingly, the evolution of the state of S *may or may not* be governed by a Hamiltonian. What the production of mixtures by measurements shows is that the interactions between S and M cannot be represented in any Hamiltonian for S. Measurements therefore do not conflict with the Schrödinger equation, since they do not satisfy the conditions of its applicability—*i.e.* do not correspond to any Hamiltonian. (Cf. Jauch [1968], p. 153.)

As Park has suggested in his [1968*a*], the impossibility of giving an account of the temporal evolution of the state of S which is independent of S's being a part of $S+M$ provides a way (different from that in section 2) of understanding Bohr's remark (cf. his [1963], p. 4) that 'the interaction between object and apparatus . . . forms an inseparable part of the [quantum] pheonomenon', which therefore has a certain 'feature of wholeness'.

5 MEASUREMENT: THE UNCERTAINTY PRINCIPLE

The principal criticism of Reichenbach which emerged in the foregoing digression on the projection postulate is that he assumes that this postulate is true of measurements generally. Much of what we said above, however, is consistent with his views. Unlike other authors we criticised, Reichenbach does not say, for example, that the Schrödinger equation requires additional terms to reduce the wave packet. He even shows himself to be aware, as we insisted above, that ascription of a wave function to a system is only meaningful relative to that system's membership in a specified assemblage (cf. Reichenbach [1944], p. 95). But the relativity of the wave function is itself a difficulty for Reichenbach, at least on van Fraassen's interpretation (cf. *above*, p. 93, footnote 2). For if the wave function of a system depends upon the reference class we view the system as a member of, then according to (1)–(3) in section 3, so do the truth-values of statements concerning its observables. Yet surely the question of whether it is true that a particle has a certain position has an answer independent of any choice of reference class.

This indicates that it is uncharitable to take (1)–(3) as an interpretation of Reichenbach—*i.e.* as equivalent to Definitions 4 and 5′. We obtain a more favourable interpretation by retreating to our original formulation of Reichenbach's view: '$u = u_i$' is true of S at time t iff a measurement of u at t yields u_i, or S is at t in an eigenstate of u (with eigenvalue u_i) prepared by a measurement of u. In this formulation, the relativity of the state is no problem, since it is made explicit that the generating conditions consist in a measurement of u.

This version of Reichenbach's theory leaves his solution to the two-slit paradox untouched. However, it requires modification of his solution to the tunneling and orbital-electron paradoxes. It is no longer possible to argue that '$E = (p_x^2/2m) + V(x)$ has the same value for every x' is indeterminate on the ground that no state is an eigenstate of both x and p_x. Since we have replaced the truth-value conditions relating to eigenstates with conditions relating to measurements, a defender of Reichenbach is now compelled to argue that it is impossible to perform (exact) measurements of simultaneous values of x and p_x. He therefore requires the usual interpretation of the uncertainty principle as forbidding such measurements.

Unfortunately, the orthodox interpretation is incorrect, though it is probably held by a large majority of writers on quantum theory. Heisenberg himself (cf. *e.g.* his [1930]) must take much of the blame for the widespread confusion. Popper in his [1935] was perhaps the first writer to penetrate the familiar obfuscations, but his work on the subject has received the attention of few quantum theorists. More recently, Park in his [1968*a*, *b*], Park and Margenau in their [1968], and Ballentine in his [1970] have succeeded almost entirely in clearing the matter up. There is no space here to repeat their arguments. We shall simply assume that they are correct in interpreting the uncertainty principle as referring to dispersions—mean deviations—in measurements upon sets of similarly prepared systems, and not to the precision of single measurements. In particular, the uncertainty principle asserts that the product of the dispersions in x and p_x is at least $\hbar/2$, but says nothing about simultaneous measurements of these quantities. (This opinion, incidentally, is expressed in a late essay by Bohr (cf. his [1963], p. 5).)

Whether arbitrarily precise measurements of x and p_x at t can in fact take place is, of course, a separate question, and one whose answer is less certain. Nevertheless, we shall assume that the examples given by Popper ([1967], pp. 26 ff.), Park and Margenau ([1968], pp. 239–42), and Ballentine ([1970], pp. 365 ff.) establish that such measurements are physically possible, and that Park and Margenau have correctly explained why some (false) versions of quantum theory contain a theorem to the contrary.

The implication for Reichenbach's quantum theory of this view of measurement, of course, is the following: while his theory seems to resolve the two-slit paradox, it cannot resolve the tunneling and orbital-electron paradoxes without utilising the false premise that simultaneous values of x and p_x cannot be measured. (This objection applies to Bohr's first theory as well.)

6 NEW PARADOXES

Thus far we have suggested that Reichenbach's theory does not provide adequate solutions to the paradoxes which motivated it. But it also runs into additional difficulties unrelated to our original three paradoxes.

The first difficulty lies in the arbitrary distinction Reichenbach is forced to make between microscopic and macroscopic objects. We noted in section 2 that Bohr sometimes held that statements about the unobserved behaviour of 'atomic objects' are meaningless. Reichenbach has objected to this that it seems arbitrary to apply such a rule to statements about unobserved atomic objects when we would refuse to apply it to statements about unobserved pieces of hardware, for example (Reichenbach [1948], p. 347). However, an analogous objection may be raised against his own theory. We do not consider statements about unobserved macroscopic objects to be indeterminate. How, then, can we justify taking this view of statements about unobserved microscopic objects? The distinction between 'macro' and 'micro' is vague and arbitrary, whereas that between 'true' and 'indeterminate' is not. How, then, can the distinctions be correlative? (Cf. Einstein [1949], p. 674.) Further, must not our physical theory ***explain*** why this difference between the large and the small exists?

In his [1948], written after his book, Reichenbach sought to answer objections of this sort. His reply was, in effect, to modify his condition for an observable's having a value (cf. p. 101, *above*) by dropping the requirement that a measurement take place. The revised condition for truth of '$u = u_i$' at t is:

(C) a measurement of u at t yields u_i, or the system is at t in an eigenstate of u (with eigenvalue u_i) with respect to some preparation.

We can see why this reply might be thought helpful by considering the case which prompted Reichenbach to make it—that of Schrödinger's cat (see Schrödinger [1935]; our treatment follows Park [1968*b*]). Suppose we incarcerate a cat in a 'hellish contraption' which will kill the cat instantly if a certain unstable atom decays. Consider the observable A whose value is 1 if the cat is alive and 0 if it is dead. The corresponding eigenstates are α and δ. ϕ is the initial unstable state of the atom and θ its state after decay. Initially, the state of atom+cat is $\Psi(0) = \alpha \otimes \phi$. By our result for correlated systems in section 4, at any later time t the state is

$$\Psi(t) = c_1(t)\alpha \otimes \phi + c_2(t)\delta \otimes \theta.$$

Our earlier result also implies that

$$\rho^t_{cat} = |\, c_1(t)\,|^2\, P_\alpha + |\, c_2(t)\,|^2\, P_\delta.$$

Hence, at any time after the preparation of these states by setting up the apparatus, the cat is in a mixture of eigenstates of A. A defender of Reichenbach might now complete his reply to Schrödinger by invoking the 'ignorance interpretation of mixtures': if a system is in a mixture of states, then it really is in one of the states, and the probability that it is in each is given by the weight which that state has in the mixture. It follows that Schrödinger's cat is either in α or δ, with probabilities $|\, c_1(t)\,|^2$ and $|\, c_2(t)\,|^2$, and (by Reichenbach's sufficient condition for an observable's having a value) that $A = 1$ or 0—*i.e.* that the cat is either alive or dead. This result holds, of course, whether or not any measurement or observation takes place to determine whether the cat is alive. We would therefore seem to have shown that Reichenbach's quantum theory, while assuming that unmeasured microscopic observables sometimes lack values, is capable of establishing, at least in this special case, that observables relating to macroscopic objects retain their values when the objects are unobserved.

In the above derivation we did not make use of the assumption that the cat is a macroscopic object and that the value of A corresponds to the truth-value of a proposition we think to be either true or false regardless of whether any measurement has taken place. It may seem, then, that a solution of this sort somehow misses the point of Schrödinger's objection. This suspicion is well-founded. First of all, the ignorance interpretation of mixtures, which this solution presupposes, leads to a contradiction. The reason is that the resolution of a mixed state into pure states is not in general unique. Park in his [1968*a*] has shown that there are cases in which $\rho = w_1\rho_1 + w_2\rho_2 = w_3\rho_3 + w_4\rho_4$, and in which the ignorance interpretation implies that some members of an ensemble are in both ρ_1 and ρ_3, where $\rho_1 \neq \rho_3$. Since the ignorance interpretation leads in some cases to two different states for a single system, and provides no way to choose between them, it is evidently unsatisfactory. So, therefore, is Reichenbach's proof that Schrödinger's cat is alive or dead.

There is a second objection to Reichenbach's attempt to reconcile his three-valued quantum logic with two-valuedness at the macroscopic level. Suppose it could somehow be established that if the cat is in a mixture of the states α and δ then it is either alive or dead. Then the objection would re-arise for the composite system atom+cat. As we saw, its state at t is given by

$$\Psi(t) = c_1(t)\alpha \otimes \phi + c_2(t)\delta \otimes \theta$$

or by its density matrix $P_{\Psi(t)}$. This state is evidently not an eigenstate of the observable B, whose value in a system S is 1 if S consists of an unstable atom and a live cat, and 0 if the atom has decayed and the cat is dead. Since it is a pure state, $\Psi(t)$ is obviously not a mixture of eigenstates of B. Reichenbach's sufficient condition for an observable's having a value therefore cannot be used to show that one or the other of these propositions is true of the composite system. Needless to say, it would not help to claim, as Heisenberg does in his [1958], p. 178, that a system as complex as a cat is always in a mixture and never a pure state like ϕ, as assumed above. This supposition would lead to a final state which was a mixture of superpositions of the type $\Psi(t)$, and not to a mixture of eigenfunctions of B. Nor, of course, would it help to view the atom+cat as parts of some larger composite system, since the same objection could be raised for the latter system, whatever it is.

Such writers as Feyerabend (cf. his [1957], p. 123) and Jauch (cf. his [1968], ch. 11) have argued that while the final state of atom+cat is pure, the difference between it and the mixture

$$\rho' = |\, c_1(t) \,|^2 P_{\alpha\otimes\phi} + |\, c_2(t) \,|^2 P_{\delta\otimes\theta}$$

could not be detected by any physically possible set of experiments. Whatever one might think of the arguments offered for this assertion, and even though it may be true, it is not relevant in the present context. In order to utilise it, a defender of Reichenbach would have to weaken his sufficient condition for an observable's having a value to require only an approximate eigenstate. But this condition is vague, and it is not clear how the vagueness could be removed in any non-arbitrary manner. How could the sharp distinction between existence and non-existence of an observable's value correspond to the diffuse distinction between states which are approximate eigenstates and those which are not? (Cf. Putnam [1965].) Even if this difficulty could be evaded, the non-uniqueness of the resolution of mixtures into pure states would obstruct any argument from an approximate mixture to an approximate eigenstate.

It is important to appreciate that the Schrödinger cat paradox does not constitute an objection to the quantum theory itself. One could object to the quantum theory on the ground that it implies that the final state of atom+cat is pure rather than mixed only if one could produce mortality statistics for large numbers of cats in contraptions. If these showed that the statistical predictions derived from

$$\rho = P_{\psi(t)}$$

are incorrect and that those derived from

$$\rho' = |\, c_1(t) \,|^2 P_{\alpha\otimes\phi} + |\, c_2(t) \,|^2 P_{\beta\otimes\theta}$$

are correct, we would have reason to reject quantum theory. In the absence of such statistics, we have every reason to assume that the statement '$\rho = P_{\psi(t)}$', since it follows from a well-confirmed theory, is true. (Cf. the apparently different attitude in Wigner [1963].)

Our final objection to Reichenbach's theory is that even if condition *C* somehow enabled him to avoid objections like Schrödinger's, it would lead to yet another paradox: one similar to that of Einstein, Podolsky, and Rosen (see their [1935]). These authors (EPR) were attempting to refute the claim that quantum theory is 'complete'. By this they meant the claim that a system has a property if and only if the state given by quantum theory specifies (with probability 1) that it does. Hence, the completeness claim is roughly equivalent to Reichenbach's condition *C* for an observable's having a value. EPR postulate that a 'sufficient condition for the reality of a physical quantity is the possibility of predicting it with certainty, without disturbing the system'. From this they were able to deduce that two conjugate quantities can have simultaneous values ('reality'). As we noted in section 2, Bohr's reply was to reject EPR's sufficient condition for the reality of a physical quantity, and to require an eigenstate with eigenvalue u_i as a necessary, and presumably also sufficient, condition for '$u = u_i$'. Now condition C also recognises the system's being in an eigenstate of u (with eigenvalue u_i) as a sufficient condition for the truth of '$u = u_i$'. By showing that this condition, though stronger than EPR's, leads to a paradox similar to theirs, we can refute both Reichenbach's reply to Schrödinger, and Bohr's reply to EPR.

For simplicity, we will consider spin rather than position and momentum (cf. Bohm [1951], ch. 22; this version also owes something to Reichenbach [1944]). Suppose that a system composed of two spin-$\frac{1}{2}$ particles is in a state with a total spin of zero and wave function

$$\Psi = \frac{1}{\sqrt{2}} a^1 \otimes b^2 - \frac{1}{\sqrt{2}} b^1 \otimes a^2,$$

where the a's and b's are eigenfunctions of spin ($\sigma_z a = a$, $\sigma_z b = -b$), and the superscripts distinguish the two particles. With respect to the preparation of this state, the first particle is evidently not in an eigenstate of σ_z^1, since the probability is $\frac{1}{2}$ that a measurement of σ_z^1 would yield 1 (or -1). Before the particles interact with anything, no other preparation has occurred with respect to which particle 1 is in an eigenstate. On Bohr's or Reichenbach's view, then, there is no number which is the value of σ_z^1. Suppose that we now measure σ_z^2 and obtain the result -1. Relative to the preparation which consists of preparing the composite state Ψ and then obtaining a value of -1 for σ_z^2, the probability is 1 that a measurement of σ_z^1 will yield the result 1. Accordingly, particle 1 is now in the eigenstate a^1 with respect to this preparation. It follows from condition C that after the measurement of σ_z^2, σ_z^1 has the value 1. Even if particles 1 and 2 are separated by light years, then, a measurement upon 2 results instantly in an event at particle 1: the acquisition of a value of σ_z^1. This result ('the modified EPR paradox') is implausible in the extreme, and conflicts with the theory of relativity.[1]

It is not paradoxical, of course, to suppose that acquisition of information affects the probabilities of distant events. If my car is destroyed in Chicago, the probability that it will be destroyed in Los Angeles is instantly reduced to zero. Plainly, this is not the instantaneous causation of a distant event. Similarly, a distant event may instantly bring it about that a system is in an eigenstate with respect to some preparation; and this is in itself not a violation of special relativity. But if we think of the wave function, not only as giving probabilities, but also as determining whether individual systems have certain physical properties, then distant events may result not only in changes in probabilities, but also in changes in physical properties—*i.e.* in events. It is the attempt to force the wave function into this dual role which results in both the Schrödinger and EPR-type paradoxes. Popper was surely right in terming this attempt 'the great quantum muddle' (see his [1967], p. 19).

---

[1] It is the fact that the value of either σ_z^1 or σ_x^1 can be determined in this way (*v.* Bohm) that enabled EPR to claim that both values exist, in accordance with their sufficient condition for the reality of a physical quantity. Bohr's requirement of an eigenstate does not permit this conclusion, of course, since there are no eigenstates of both σ_z^1 and σ_x^1. But this requirement does result in the above causal anomaly.

The modified EPR paradox is not an effective argument against the first Copenhagen view (section 2), according to which the observable σ_z^1 does not have a value until particle 1 itself interacts with a measuring apparatus. Condition *C* begs the question against that first view by assuming that an observable may have a value before measurement if the system is in an eigenstate. The modified EPR paradox is of interest only when taken in conjunction with a paradox like Schrödinger's. The case of the cat shows that the requirement that an observable be measured is too strong a condition on its having a value. What the new paradox shows is that the attempt to meet Schrödinger's objection by weakening this requirement to that embodied in condition *C* is objectionable in its own right.

The problem of specifying the condition under which an observable has a value therefore remains unsolved as this article ends. Again, however, we have found no objection to the quantum theory proper—as axiomatised, for example, in Mackey [1963]. For this theory concerns statistical predictions of the results of measurements, and does not itself embody any attempt to solve the general problem of when an observable has a value.[1] Both the Schrödinger cat paradox and the original and modified EPR paradoxes should rather be read as objections to Bohr's and Reichenbach's attempts to adjoin to the standard quantum theory certain conditions for an observable's having a value and to embody these in deviant logics.

7 CONCLUSION: BOHR RECONSIDERED

Since writers of so many different persuasions have been able to find support for their views in Bohr's sometimes oracular pronouncements, it is not altogether surprising that the minimal statistical interpretation advocated here can be found in his writings, in addition to the two distinct Copenhagen interpretations discussed in section 2.

At times Bohr adopts what looks like the view discussed in the philosophical literature under the heading 'instrumentalism': 'the quantum-mechanical formalism represents a purely symbolic scheme permitting only predictions . . . as to results obtainable under conditions specified by means of classical concepts' (Bohr [1958], p. 40). In general, he says, 'when speaking of a conceptual framework, we refer merely to the unambiguous logical representation of relations between experiences' or, as he says one paragraph later, 'between measurements' (*ibid.*, p. 68). These passages suggest that those well-formed formulas of quantum theory—*e.g.* the Schrödinger equation—which are not themselves statements of measurement-frequencies have a syntax but no semantics. That is, they have formation rules and transformation rules which permit the derivation of genuine statements (of measurement-frequencies), but are not themselves genuine statements—*i.e.* bearers of truth-values.

---

[1] It would be a mistake, then, to interpret EPR as claiming that the correct state of the composite system is not the pure state $\Psi = \frac{1}{\sqrt{2}} a^1 \otimes b^2 — \frac{1}{\sqrt{2}} b^1 \otimes a^2$ but the mixture $\rho = \frac{1}{2}(P_{a^1 \otimes b^2} + P_{b^1 \otimes a^2})$. This common misinterpretation (cf. Furry [1936]; Jauch [1968]) reads the original EPR paradox as an objection to the quantum theory, according to which the correct statistical predictions are given by Ψ and not ρ.

It is hard to know what plausible arguments Bohr could offer against the 'naive' reply that '$\Psi(t) = \mathrm{e}^{-itH}\Psi(0)$' has an interpretation—*sp.*, as a statement about how wave functions change in time—and is true or false according to whether they really change that way. Perhaps, then, we arrive at a more sympathetic interpretation of Bohr if we ignore the instrumentalist cast of his assertions, and read him not as denying that 'Ψ' has an interpretation, but simply as denying that 'Ψ' is to be interpreted as determining whether an individual system possesses or lacks certain physical magnitudes, as claimed by Reichenbach and by Bohr's other self discussed in section 2. Ψ's role, rather, is simply to determine the probabilities of various measurement-results. Just as we have claimed that the adoption of this austere minimal statistical interpretation allows one to avoid the Schrödinger and EPR-type paradoxes, so Bohr claims that if we interpret the quantum theory as referring to no 'phenomena' except (frequencies of) results of observations, 'the observational problem in quantum physics is deprived of any special intricacy' (*ibid.*, p. 73). He also suggests (*ibid.*) that this austere position, which is agnostic about unmeasured magnitudes, enables one to avoid being compelled to speak of 'disturbance of phenomena by observation' or 'creation of physical attributes to atomic objects by measurements'. (As is shown elsewhere (Gardner [1972]), this is the fate of Popper and others who assert that all physical magnitudes always have values.) Finally, Bohr expresses agreement with our view that an advantage of the minimal statistical interpretation is that 'all departures from common language and ordinary logic are entirely avoided' (Bohr [1963], p. 6).

Thus does the study of Bohr combine the pleasures of patricide with those of ancestor-worship.

Harvard University

REFERENCES

Ballentine, L. E. [1970]: 'The Statistical Interpretation of Quantum Mechanics', *Review of Modern Physics*, **42**, pp. 358–81.

Bohm, D. [1951]: *Quantum Theory*.

Bohm, D. and Bub, J. [1966]: 'A Proposed Solution of the Measurement Problem in Quantum Mechanics by a Hidden Variable Theory', *Review of Modern Physics*, **38**, pp. 453–69.

Bohr, N. [1935]: 'Can Quantum Mechanical Description of Physical Reality be Considered Complete?', *Physical Review*, **48**, pp. 696–702.

Bohr, N. [1958]: *Atomic Physics and Human Knowledge*.

Bohr, N. [1963]: *Essays 1958–1962 on Atomic Physics and Human Knowledge*.

Einstein, A., Podolsky, B. and Rosen, N. [1935]: 'Can Quantum Mechanical Description of Reality be Considered Complete?', *Physical Review*, **47**, pp. 777–80.

Einstein, A. [1949]: 'Reply to Criticisms', in P. A. Schilpp (*ed.*): *Albert Einstein: Philosopher-Scientist*, pp. 665–88.

Feyerabend, P. K. [1957]: 'On the Quantum Theory of Measurement', in S. Körner (*ed.*): *Observation and Interpretation*, pp. 121–30.

Fraassen, Bas C. van: *The Labyrinth of Quantum Logics*. (Unpublished.)

Furry, W. H. [1936]: 'Note on the Quantum-Mechanical Theory of Measurement', *Physical Review*, **49**, pp. 393–9.

Gardner, M. R. [1971]: 'Is Quantum Logic Really Logic?', *Philosophy of Science*, **38**, pp. 508–29.

Gardner, M. R. [1972]: 'Quantum-Theoretic Realism: Kochen and Specker *v.* Popper and Einstein', *British Journal for the Philosophy of Science*, **23**, pp. 13–23.

Heisenberg, W. [1930]: *Physical Principles of the Quantum Theory*.

Heisenberg, W. [1958]: *Physics and Philosophy*.

Jauch, J. M. [1968]: *Foundations of Quantum Mechanics*.

Mackey, G. [1963]: *Mathematical Foundations of Quantum Mechanics*.

Neumann, John von [1932]: *Mathematical Foundations of Quantum Mechanics*.

Park, J. L. [1968*a*]: 'Nature of Quantum States', *American Journal of Physics*, **36**, pp. 211–26.

Park, J. L. [1968*b*]: 'Quantum Theoretical Concepts of Measurement', *Philosophy of Science*, **35**, pp. 205–31, 389–411.

Park, J. L. and Margenau, H. [1968]: 'Simultaneous Measurability in Quantum Theory', *International Journal of Theoretical Physics*, **1**, pp. 205–83.

Popper, K. R. [1935]: *Die Logik der Forschung*.

Popper, K. R. [1967]: 'Quantum Mechanics without "the Observer" ', in M. Bunge (*ed.*): *Quantum Theory and Reality*, pp. 7-44.

Putnam, H. [1965]: 'A Philosopher Looks at Quantum Mechanics', in R. Colodny (*ed.*): *Beyond the Edge of Certainty*, pp. 75–101.

Putnam, H. [1970]: 'Is Logic Empirical?', in R. S. Cohen and M. Wartofsky (*eds.*): *Boston Studies in the Philosophy of Science*, **5**, pp. 216–41.

Reichenbach, H. [1944]: *Philosophic Foundations of Quantum Mechanics*.

Reichenbach, H. [1948]: 'The Principle of Anomaly in Quantum Mechanics', *Dialectica*, **2**, pp. 337–48. Reprinted in H. Feigl and M. Brodbeck (*eds.*): *Readings in the Philosophy of Science*, 1953, pp. 509–20.

Schrödinger, E. [1935]: 'Die gegenwärtige Situation in der Quantenmechanik', *Naturwissenschaften*, **23**, pp. 807–12.

Shimony, A. [1963]: 'The Role of the Observer in Quantum Theory', *American Journal of Physics*, **31**, pp. 755–73.

Wigner, E. P. [1963]: 'The Problem of Measurement', *American Journal of Physics*, **31**, pp. 6–15.

Brit. J. Phil. Sci. **23** (1972), 111–116 *Printed in Great Britain*

The Incompatibility of Mach's Principle and the Principle of Equivalence in Current Gravitation Theory

by JAMES F. WOODWARD and WOLFGANG YOURGRAU

Both Mach's Principle and the Principle of Equivalence have been extensively discussed in the recent past. However, the compatibility of the Principle of Equivalence with Mach's Principle is a question which, for some time now, has received little explicit attention. To the present it has been assumed that if gravitational field equations could be found which globally yield no solution in an empty universe, Mach's Principle would be subsumed under these equations. (This was the opinion of Einstein which has widely prevailed to this day.) The object of this paper is to show that for the two principles to be compatible this condition on the field equations is merely a *necessary*, but not a *sufficient* condition and that, given accepted theories of gravitation, the two principles are incompatible.

Mach's and the Equivalence Principles are well known, but, as neither has a succinct and generally held definition, let us now give definitions of these principles that will suffice for our purposes in the following discussion. Mach's Principle may be stated: the global distribution of matter in the universe (Mach's 'fixed stars') causally determines *by itself* the properties of the local inertial field and thus the inertial properties of local bodies and the local inertial frame of reference.[1] Note that we say *the* rather than *a* local inertial frame of reference. We make this distinction because recent criticism indicates convincingly that globally the principle of general relativity (as formulated by Einstein) does not hold. Specifically, the failure stems from the fact that there exists only one frame of reference at any given point in the universe in which the cosmic blackbody radiation appears isotropic.[2] However, the unique frame of reference in which the universe appears isotropic with respect to *all* phenomena is that maintained by Mach to be at rest with respect to the local inertial field as defined by the 'fixed' stars. It is firmly established that Mach's Principle is empirically verified with significant accuracy.

The Principle of Equivalence simply asserts that the effects of acceleration and gravitation are physically indistinguishable, or, that inertial and

---

[1] Cf. Mach [1883], chapter II, section 6.

[2] Cf. Bergmann [1970].

Received 16 *August* 1971

passive gravitational mass are identically the same. We should, however, make a distinction here between the so-called 'strong' and 'weak' forms of the Equivalence Principle. We take the 'strong' form of the Equivalence Principle to be the claim that, at least in a sufficiently small region of space-time, a gravitational field is in all respects identical to and indistinguishable from an accelerated frame of reference. The 'weak' form consists of the assertion that passive gravitational and inertial *rest-mass* are quantitatively equal. It is this, the 'weak' form, which is tested by Eötvös-type experiments. These two forms of the Equivalence Principle are significantly different. Whereas the weak form of the Equivalence Principle involves only rest-mass, the strong form extends this statement to bodies with zero rest-mass—specifically, the electromagnetic field. Einstein, of course, realised that this extension constitutes an independent assumption as he explicitly included electromagnetic radiation in his definition of matter.[1]

Actually, the above distinction may be quite important as Duchesne and Vigier have recently presented empirical evidence which indicates that the photon has a finite rest-mass.[2] If, in fact, this is so, then the Principle of Equivalence must be subjected to close re-examination. Wheelon has shown that if the photon has a finite rest-mass, then the deflection of light by the solar gravitational field must be a function of the frequency of the light.[3] Woodward and Yourgrau have shown further that if the deflection of light in a gravitational field is frequency-dependent, then one may always, at least in principle, distinguish between a gravitational field and an accelerated frame of reference.[4] Although this argument forces one to abandon the strong form of the Equivalence Principle, it does not vitiate the weak form of the Equivalence Principle.

On aesthetic grounds Mach's Principle is also very appealing—in our opinion compelling—by reason of its being an immediate consequence of the most general statement of the principle of relativity: for a single body in an otherwise empty universe *no* state of motion, and hence acceleration (translational or angular), can be specified.[5] Einstein himself, for obvious reasons, regarded Mach's Principle as a profound physical truth. He attempted, in vain, to incorporate it in his general theory of relativity (of 1917) by the introduction of the famous cosmological constant term $(\lambda g_{\mu\nu})$ in his gravitational field equations. It was Einstein's hope that by inserting this term in the field equations, they would have no solution for a massless universe, and an empty universe would therefore be physically meaningless. Einstein did not succeed in this attempt as de Sitter showed within a year by demonstrating that a massless, expanding universe was consistent with Einstein's field equations.[1]

---

[1] Cf. his [1916], section 14.

[2] Cf. their forthcoming paper in *Physical Review Letters*.

[3] Cf. his [1952].

[4] Cf. their [1970].

[5] Koyré seems to have missed this elementary point. Cf. his [1950], p. 298.

We have stated that a null solution of the gravitational field equations in an empty universe is a necessary, but not a sufficient condition for Mach's Principle and the Principle of Equivalence to be compatible. What, then, is the both necessary and sufficient condition for compatibility? This condition may be shown to be that the gravitational field equations—and hence the inertial field equations—must be null for one body in an otherwise empty universe. This state of affairs follows because, according to Mach's Principle, one cannot define, for such a body, a state of motion; therefore, since the Principle of Equivalence demands that dynamical motion may be treated as identically the same as the influence of gravitation, no gravitational field can be uniquely specified.

Let us demonstrate, in a different way, the validity of the claim that the inertial field for one body alone in the universe must be null by the following *Gedankenexperiment.* Consider a body in a universe with a finite mass-density—that is to say, a massive body in a universe like ours. We now, by some unspecified means, isolate the body under consideration from the influence of all physical fields (inertial, gravitational, electromagnetic, *etc.*) produced by *all* of the external matter in the universe. A body so isolated will, under Mach's Principle, have no inertia. This follows because Mach's Principle demands that the local inertial field is produced by the rest of the matter in the universe. Since we, by isolating our body from all external influences, have effectively obliterated the local inertial field, clearly, our body is now inertialess. The immediate consequence of the body being inertialess is that, relative to an external observer, this body may possess any arbitrary state of motion—*accelerated* as well as kinematic—without the application of an external force. Such a body would thus be equivalent to a body in an otherwise empty universe since they are identical in so far as they both may have any arbitrary state of motion (and this is, of course, *the* distinguishing characteristic of a body in a vacuous universe according to Mach's Principle).

It is reasonable to assume that if our body were to possess an inertial field specifiable at all points in the space-time continuum, this field would act upon the external matter in the universe. But this would be a violation of the law of action and reaction (since the isolated body would act on others without being acted upon itself) and the first law of thermodynamics (as, in any system containing such a body, energy would not be conserved). We therefore conclude that our isolated body, and hence a body in a vacuous universe, can only possess a null inertial field. Thus, if either form of the Principle of Equivalence is valid, our body cannot possess either an inertial or a gravitational field.[1]

---

[1] Cf. North [1965], pp. 83–92.

Under the currently held theories of gravitation (Newton's and Einstein's), both of which are viewed on the basis of the field concept introduced by Maxwell (where the field is a separate and real physical entity), the demand that for one body no gravitational field exist is absurd. Thus, given the accepted theories of gravitation, we may infer that Mach's Principle and either form of the Principle of Equivalence are simply incompatible, *i.e.* one of them must be wrong—or the theories of gravitation are wrong—or both the theories of gravitation and one of the principles are wrong. If Duchesne and Vigier's work, noted above, is accurate, then at least the strong form of the Equivalence Principle is incorrect. But if the Principle of Equivalence in its strong form is vitiated, then so too is the theory of gravitation currently believed to be correct (Einstein's).

It appears to us, given the choice between Mach's Principle and the Principle of Equivalence, one would have to opt for Mach's Principle—particularly in view of Duchesne and Vigier's work on the photon rest-mass. The claim that one can make no statement about the inertial properties (or, for that matter, most other properties) of one body alone in the universe appears so intuitively obvious that it is practically beyond doubt. On the other hand, although Mach's Principle implies that the inertial properties of a body are determined by the gravitational fields produced by the external matter in the universe, this is not equivalent to the demand that inertia and gravitation are identically the same as required by the strong (but not the weak) form of the Equivalence Principle. To our way of thinking, such a claim is by no means obvious. In fact, there is empirical evidence that, at least for the case of the interaction of the electromagnetic and gravitational fields, this claim is unwarranted.[2]

In searching for a solution to this dilemma one thing seems to be clear: nothing like Newtonian or Einsteinian gravitation, where one can specify the time-independent gravitational field of a body in the absence of all other matter, will work. The specifiability of the field for the one body case is a consequence of the fact that in both Newtonian and Einsteinian gravitation the monopole component of the field is static; that is, the monopole component of the field is not radiational. Therefore, this component of the field exists (and persists indefinitely) even if only one body exists in the universe. But, on the contrary, if the monopole component of the field were radiational, no difficulty would arise as a single body would rapidly radiate away its gravitational field. Once this had happened it would satisfy both Mach's and the 'weak' Equivalence Principles as it could have no inertial interaction with its now essentially nonexistent field. Inertia, in the light of such a theory, would simply be the measure of the resistance of a body to changes in its state of motion with respect to the background gravitational field produced by the rest of the mass in the universe—exactly as implied by Mach's Principle. The mean value of the inertial field in a universe with a finite mass-density would be quantitatively related to the point at which dynamic equilibrium between radiation and absorption were achieved by the gravitational field. (Thus, *if* the universe were expanding and consequently the mass-density were decreasing, the mass of all bodies and all physical constants which depend on mass would be decreasing in time—as in the theories of Dirac, Jordan, and Brans and Dicke.)

---

[1] Here we have isolated the body from all external fields. Even if one does not subscribe to the strong form of the Principle of Equivalence, one may assume that the inertial and gravitational fields are causally connected—that the gravitational field is the source of the inertial field as implied by the weak form of the Equivalence Principle. Hence, apparently it would suffice to isolate a body only from the gravitational fields of external matter to render it inertialess.

[2] Cf. a forthcoming article by the authors in *Nuovo Cimento*.

Naturally, theories of gravitation embodying the characteristics described above are not unknown. They were devised before Mach, and several have been devised since. To name but a few, there are those of Lesage,[1] C. F. Brush (a rather grotesque attempt),[2] and Quirino Majorana (a much more plausible attempt).[3] The most interesting, however, is that recently proposed by Freund, *et al.*[4] For reasons other than those discussed here, they have tried to show that one must take a Yukawa-type potential for the gravitational field. Such a potential yields a field with finite rest-mass transfer particles which are propagated with a finite velocity and, hence, is a radiation field. Such a solution to the dilemma would be appealing on other, aesthetic grounds. If, as indicated by Duchesne and Vigier, the photon has a finite rest-mass, then the electromagnetic field may also be a Yukawa-type, finite range field. Thus, all known physical fields of force would be finite range fields with finite rest-mass transfer particles propagated at finite velocities. Aesthetically, this consistency would seem desirable.

Be all this as it may, it is apparent that empirical research on the nature of the gravitational field is definitely required. Hopefully such research will show whether Mach's Principle, the Principle of Equivalence, or current theories of gravitation are wrong, for it is impossible for all of them to be correct.

[1] Cf. his [1749]. [2] Cf. his [1929]. [3] Cf. his [1920]. [4] Cf. their [1969].

JAMES F. WOODWARD
WOLFGANG YOURGRAU
Department of History and Philosophy of Science
University of Denver

REFERENCES

Bergmann, P. G. [1970]: 'Cosmology as a Science', *Foundations of Physics*, **1**, pp. 17–21.

Brush, C. F. [1929]: 'Kinetic Theory of Gravitation', *Proceedings of the American Philosophical Society*, **68**, pp. 55–68.

Einstein, A. [1916]: 'Die Grundlage der allgemeinen Relativätstheorie', *Annalen der Physik*, **49**, pp. 769–822. Reprinted in English in A. Sommerfeld (*ed.*): *The Principle of Relativity*, 1923, pp. 109–64.

Freund, P. G. O., Maheshwari, A. and Schonberg, E. [1969]: 'Finite-Range Gravitation', *Astrophysical Journal*, **157**, pp. 857–68.

Koyré, A. [1950]: 'The Significance of the Newtonian Synthesis', *Archives Internationales d'Histoire des Sciences*, **3**, pp. 291–311. Reprinted in A. Koyré *Newtonian Studies*, 1965, pp. 3–24.

Lesage, G. L. [1749]: *Essai sur l'origine des forces mortes* (manuscript in the library of the University of Geneva).

Mach, E. [1883]: *Die Mechanik in ihrer Entwicklung historisch-kritisch dargestellt* (Sixth English edition: *The Science of Mechanics*, 1960).

Marjorana, Q. [1920]: 'Theoretical and Experimental Researches on Gravitation', *Philosophical Magazine*, **39**, pp. 488–504.

North, J. D. [1965]: *The Measure of the Universe*.

Wheelon, A. D. [1952]: 'Gravitational Deflection of Photons with Non-vanishing Rest Mass', *Physical Review*, **85**, pp. 383–84.

Woodward, J. F. and Yourgrau, W. [1970]: 'A Paradox in the Interaction of Gravitational and Electromagnetic Fields', *Nature*, **226**, pp. 619–21.

Discussions

ON THE MACH PRINCIPLE AND RELATIVE SPACE-TIME

THE purpose of this note is to discuss what I feel to be a common misconception among physicists and philosophers of science on what it is that is asserted by the Mach principle, on the one hand, and on the other, what are its implications. If one should stay close to Mach's original discussion in his [1883], this principle (named later by Einstein) has to do with Mach's interpretation of inertia—that *physical manifestation* of interacting matter that is the cause of its resistance to a change of state of rest or constant velocity. The principle itself does not *directly* relate to space and time. It is the assertion that the inertia of any quantity of matter is a measure of its dynamical coupling with all of the other constituent matter within the considered closed system (*in principle*, the universe).

The implication of this principle is that any physical system is necessarily *closed*, since with this view *no matter* can be considered as free of the remaining components of the system. The latter conclusion appealed to Mach's anti-metaphysical stand since it rejected the assumption that the whole is made out of 'free' bits of matter that, in principle, could interact, but equally could be free to enter the unobservable states of noninteraction. Nevertheless, it is still true that this view of matter in terms of a *fundamentally closed system* need not be motivated by the positivistic philosophy. The view of a closed system, without actual parts, is also consistent with the continuous field concept (with the philosophic stand of realism) that is implicit in the theory of general relativity.

It is well known that Mach's interpretation of inertia strongly influenced Einstein's development of relativity theory. It is interesting that even though Mach's view was motivated by his positivistic philosophic stand, it led Einstein to the stand of realism—the 'underlying reality' being the closed system, without parts, necessarily described in his theory with the continuous field concept.[1] It is unfortunate that many mistakingly view the Mach principle as a unique consequence of the philosophy of positivism, perhaps because of the tendency to assign a single label to everything written by a particular scholar, rather than sifting portions of the man's philosophy that might have more scientific validity than other portions. (For example, one should not say that because of Kepler's astrological writings, his law of equal areas for planetary motion is to be rejected as mystical!)

Since the Mach principle is not more than an assertion about a *physical manifestation* (inertia) of interacting matter, it does not in itself imply relational properties between spatial points. Of course, should one accept the concept of absolute space at the outset, then inertia could be interpreted in terms of a relation between a free quantity of matter and this absolute space, without considering other bodies. Inertia could also be interpreted as an intrinsic feature of a free body (atom) that moves about in the absolute space, that is a measure of its *potential* to resist external forces, but is in itself unrelated to the space. Mach's positivistic stand forced him to reject both of these alternatives (absolute space and atomism) since neither, by themselves, directly relate to the human being's sensations. Thus the positivistic aspect of his philosophy led him to interpret the inertia of matter in terms of its dynamical coupling with all other matter.

---

[1] In a recent publication (Sachs [1970]) I have tried to analyse the influence of different *implicit* aspects of Mach's philosophy on contemporary physics.

But also in accordance with Mach's conclusion about the inertia of matter in terms of interaction, as well as his rejection of absolute space and atomism, came the theory of general relativity—a theory based on a relational interpretation of space and time in terms of the continuous field concept. While the Mach principle deals only with matter and does not directly imply anything about space and time, general relativity does relate the material content of the closed system to the geometry of space-time. The empirical evidence for the relational doctrine of space-time should then follow from observations that would support the *full form* of the theory of general relativity—a form that incorporates the Mach principle.[1, 2]

A second point that I should like to make has to do with what appears to me as a prevailing confusion between *one of the implications* of the Mach principle—the influence of the distant stars on the local properties of matter—and the principle itself. Indeed, according to this principle, distant stars must influence the local properties of matter in terms of its inertia. But it could very well be that the strength of coupling between the stars (*i.e.* the rest of the universe) and, say, an electron observed on Earth, is sufficiently weak that one could safely neglect its contribution to the electron's inertia—yet still maintaining the validity of the Mach principle! In this case, a good mathematical approximation in deriving the electron mass from the closed system, might be to consider as closed only the electron and its immediate environment—the 'physical vacuum' (involving electron-positron pairs and electromagnetic radiation) that we find necessary to explain features of elementary particle and atomic physics, as well as the remaining background matter in which one might study the electron's physical properties. (The observed isotropy and homogeneity of the electron's inertia would imply that the primary dynamical coupling that leads to its mass is in terms of the field properties of space-time, according to general relativity theory, as determined by the 'physical vacuum'.)

The main point here is the contention that the influence of the stars on the inertia of a local quantity of matter is only one of the implications of the Mach principle—it is not the principle itself! The principle could be found to be perfectly valid without ever having to evoke empirical astronomical evidence as a crucial test. On this point, Mach made the following comment in his [1883]: '(the) objection, that a comparison of masses satisfying my definition can only be effected by astronomical means, I am unable to admit. Masses produce in

---

[1] I have recently published the results of a theoretical analysis that is in the direction of supporting such a theory. (Sachs [1968].)

[2] A recent interesting exposition of views in contrast with these comments on the role of the Mach principle *vis à vis* the relational doctrine of space-time were presented in Hooker's [1971].

each other accelerations in impact, as well as subject to electric and magnetic forces, and when connected by a string in Atwood's machine.'

MENDEL SACHS
Department of Physics
State University of New York at Buffalo

REFERENCES

Hooker, C. A. [1971]: 'The Relational Doctrines of Space and Time', *British Journal for the Philosophy of Science*, **22**, pp. 97–130.

Mach, E. [1883]: *Die Mechanik in ihrer Entwicklung historisch-kritisch dargestellt* (Sixth English edition: *The Science of Mechanics*, 1960).

Sachs, M. [1968]: 'On Positive-Definiteness of Inertial Mass, Nonpositive Definiteness of Electromagnetic Coupling and the Mach Principle from General Relativity', *Nuovo Cimento*, *53B*, pp. 398–414.

Sachs, M. [1970]: 'Positivism, Realism and Existentialism in Mach's Influence on Contemporary Physics', *Philosophy and Phenomenological Research*, **30**, pp. 403–420.

RELATIVITY AND THE REALITY OF PAST AND FUTURE EVENTS

In his [1967] Hilary Putnam concludes that all events in special relativistic space-time, whether past, present, or future, are equally real, *i.e.* that a tenseless concept of existence is the appropriate concept of existence in a special relativistic world. Although I believe this conclusion is correct, I think Putnam's argument for it is not. Therefore, in this paper I will first criticise one of Putnam's main assumptions and then I will sketch a new argument for the reality of all events in special relativistic space-time that escapes this criticism.

In his argument Putnam assumes that if me-now is at a space-time point P, then of the events outside the lightcone of P and with respect to my frame of reference, those simultaneous to me-now are in my present, those later than me-now are in my future, and those earlier than me-now are in my past. However, as against this, note that in special relativity, temporal relations between events at P and events outside of P's lightcone are conventional,[1] depending not only on the frame of reference chosen but also on the *definition of simultaneity adopted.* Being merely conventional such temporal relations cannot have physical significance and thus, such concepts as past, present, and future are not physically significant when, with respect to P, they are applied to events outside of P's lightcone. This ties in also with the fact that if we adopt the usual definition of distant simultaneity of $\epsilon = \frac{1}{2}$ in every frame of reference, the temporal relations between events at P and events outside P's lightcone are not relativistically invariant (*i.e.* the same in every reference frame). And we would expect physically significant concepts of past, present, and future to be relativistically

---

[1] Depending on the choice of ϵ, $0<\epsilon<1$, along a line in space, the speed of light will be $\breve{c} = c/2\epsilon$ in one direction and $\bar{c} = c/2(1-\epsilon)$ in the opposite direction. The average round trip speed of light travelling a distance d in one direction and a distance d back again along the line is

$$\frac{2d}{d/\breve{c}+d/\bar{c}} = \frac{2}{2\epsilon/c+2(1-\epsilon)/c} = c, \text{ and so it is independent of } \epsilon.$$

invariant ones so that the past, present and future of an event at P are the same in every frame of reference.

Although in the above criticism of Putnam's assumption I have emphasised the conventionality of distant simultaneity, I think this fact can also be used to give an argument in support of Putnam's conclusion that all events in special relativistic space-time are equally real. In what follows I will sketch such an argument.

Consider a space-time point P with its associated lightcones. We know that to set up a coordinate system at P we must not only adopt an inertial frame of reference but also a definition of distant simultaneity. As is well known, we are free to choose ϵ in such a way that any *given* event outside of P's lightcone is simultaneous to an event at P in the chosen frame of reference. So any event outside of P's lightcone *can be considered* simultaneous to an event at P. Now let us assume we are accepting, with Putnam, as a common sense belief, that all things (events) that exist now are real. Assuming I-now is real and at P, I should conclude that since I can consider it to be occurring now, any event outside my lightcone can be considered real. But while distant simultaneity is a matter of convention, being real, I take it, cannot be merely a matter of convention. Thus, if an event can be considered real it must be real and so all events outside the lightcone of P (of me-now) are real.

It might be objected here that the belief that all things that exist now are real is true only for an absolute now (or present) and thus does not apply, in special relativity, to events spatially separated from P. So we cannot conclude that since an event can be considered present (*i.e.* can be made simultaneous to an event at P by a suitable choice of ε) it can be considered real. According to this objection, the appropriate relativistic analogue of our ordinary or Newtonian concept of the present of the events at P is just the events at P, for these are the only events that are absolutely simultaneous to the events at P.[1]

To meet this objection note that while in Newtonian space-time there is no difference between the class of events simultaneous to an event at P and the class of events that can be considered simultaneous to an event at P, this is not true of special relativistic space-time. Depending on our definition of distant simultaneity, the class of events simultaneous to an event at P, with respect to some frame of reference, will be one or another subset of the class of events that can be considered simultaneous to an event at P.[2]

Now in terms of *actual physical or experimental facts*, it is the class of events that can be considered simultaneous to an event at P, and not the class of events absolutely simultaneous to the events at P, that plays the role in special relativity that the class of events simultaneous to P plays in Newtonian space-time. In each they are the class of events that are not causally connectable with P. And while the class of events simultaneous to an event at P, with respect to some frame of reference, is not a relativistic invariant, the class of events that can be considered simultaneous to events at P is such an invariant. It is just the class of events outside of P's lightcone.

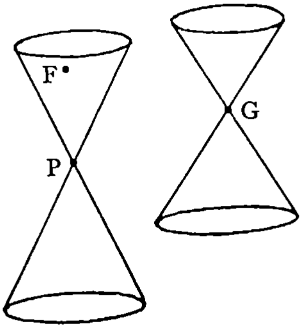
Thus, I conclude that the belief that all things that exist now (or are in the present) are real is also true for relativistic space-time, with one qualification. In special relativity the absolute present of an event at P is not the class of events simultaneous with the event with respect to some frame of reference but rather it is the class of events located outside of P's lightcone (plus the events at P). Similarly, events in the upper lobe of P's lightcone are in P's (or events at P) absolute future and events in the lower lobe are in the absolute past.

---

[1] This view is developed by Howard Stein in his [1968].

[2] Again, this latter class is the event class, any member of which is simultaneous to an event at P, with respect to any given inertial frame, on a suitable definition of ϵ.

Now given that I-now at P is real, then all the events outside of P's lightcone are real, it is easy to show that *all* events in special relativistic space-time are real. For example, for any point F we choose in the upper lobe of P's lightcone we can find a point G outside P's lightcone such that F is outside G's lightcone. Since events at G are outside of P's lightcone, these events are real since me-now at P is real. But events at F are outside G's lightcone and since events at G are real, so are the events at F. And since for any point F′ within or on P's lightcone, we can find a point G′, outside of P's lightcone; such that F′ is outside the lightcone of G′, we can conclude that all events in space-time are real.



Finally, note that in switching to G's lightcone we did not have to refer to anyone's frame of reference. Although we started earlier by referring to my frame of reference (the frame at rest with respect to I-now), the result that all the events outside of P's lightcone are in the absolute present of I-now is independent of the frame of reference or choice of ε. This is because for any space-time point P, the division of space-time points into those within P's lightcone, those on P's lightcone, and those outside of P's lightcone, is an absolute or intrinsic feature of special relativistic space-time. It is this that enabled me to conclude that if events at G are real because events at P are, the events at F are real because the events at G are, independently of any particular frames of reference.

ROBERT WEINGARD
Rutgers College

REFERENCES

Putnam, H. [1967]: 'Time and Physical Geometry', *Journal of Philosophy*, **64**, pp. 240–7.
Stein, H. [1968]: 'On the Einstein-Minkowski Space-Time', *Journal of Philosophy*, **65**, pp. 5–23.

Brit. J. Phil. Sci. 23 (1972), 123–190 *Printed in Great Britain*

Review Articles

THE LOGIC OF STATISTICAL INFERENCE[1]

To review a book seven years after its publication is unusual. The distribution of elapsed times between publication and review is probably multimodal, with a peak at a relatively short time, and subsidiary peaks at times corresponding to jubilees, centenaries, and so forth. It is a measure of the importance of Hacking's work that, in spite of the fact that the foundations of statistical inference have for ten years past been an area of very active controversy, a discussion restricted to his major theses still seems appropriate and up-to-date.

Re-reading the book one is again impressed with its easy-flowing style, full of felicitous phrases—such as 'cheerful concordat' to describe the current state of divided opinion on the foundations of set theory—but with careful attention to logical niceties. It will have been read by all who have been concerned with the foundations of statistical inference, and it is to be hoped that it will continue to be read by more and more, especially by mathematical statisticians who are all too prone to hare off into abstract mathematics without taking proper care to ensure that their mathematical model is relevant to the scientific or practical situation.

The simplest mathematical models for inferential processes are those which were first explicitly set forth by Neyman and Pearson. The elements are (i) a sample space S of possible results of the experiment in question; for instance, if we are tossing a penny ten times, S consists of the 2^{10} sequences like *HHTHTTTHTH* which could represent the results of the tosses, in the order in which they occurred; (ii) a parameter space Ω of possible values for an unknown parameter θ; for instance, Ω might consist of the points in the open unit interval $\{\theta : 0 < \theta < 1\}$, or the closed unit interval $\{\theta : 0 \leqslant \theta \leqslant 1\}$; (iii) a function $p(x,\theta)$ of two variables, x ranging over S and θ ranging over Ω, specifying the probability of getting the result x when the true value of the parameter is θ; for instance we may have

$$p(x,\theta) = \theta^{r(x)}(1-\theta)^{s(x)}$$

where $r(x)$ is the number of H's and $s(x)$ is the number of T's in x. Neyman and Pearson then characterised three types of inferential problem, (i) hypothesis testing, to 'accept' or 'reject', on the basis of x, the proposition that θ belongs to a specified subset Ω_0 of Ω; for instance, in the above example, Ω_0 might consist of the single point $\theta = \frac{1}{2}$, corresponding to the penny being unbiased; (ii) point estimation, to find a function $t(x)$, taking values in Ω which would in some sense or other be 'near' to θ; for instance we might take $r(x)/10$ as our estimate of θ in the example referred to, or perhaps $(r(x)+1)/12$; (*iii*) interval estimation, to find two functions $L(x)$ and $U(x)$, taking values in Ω, such that the probability is at least, say, 0·95 that θ lies between $L(x)$ and $U(x)$, and certain other conditions are satisfied, such as that the probability, for any point in Ω, that it lies between $L(x)$ and $U(x)$ is maximised when this point coincides with θ.

---

[1] Review of Hacking, I. [1965]: *The Logic of Statistical Inference.* Cambridge: Cambridge University Press. £2.50. Pp. ix+232.

Neyman and Pearson's models were extended by Wald with the addition of two more elements, (iv) a decision space D of possible actions to be taken, and (v) a loss function $L(d,\theta)$ measuring the loss incurred by taking action d when the parameter value is in fact θ. Aside from welcoming Wald's extended model, neither Neyman nor Pearson ever insisted that these models exhausted all the possible inferential situations that could arise; indeed, Pearson especially went out of his way on more than one occasion to stress that statistical inference could be applied to a great variety of types of situation, for which any particular mathematical model might well have extremely limited validity. In spite of this, the scope offered by the Neyman-Pearson-Wald models for purely mathematical analysis was so rich that they dominated mathematical statistics for nearly a quarter of a century; and the dominance went so far that other types of model, such as the Bayesian model (which adds, as a fourth element to the three of Neyman and Pearson a probability distribution (the 'prior') on the parameter space Ω), or the model underlying Fisher's fiducial argument, or the compound decision model of Robbins, were rejected out of hand or, at best, ignored by many mathematical statisticians. Thus, for example, although it was pointed out in 1943 by Molina, and again in 1946 by the present writer, that the Bayesian model may often be more appropriate to problems of industrial sampling inspection than is the Neyman-Pearson hypothesis testing model, most writers on the subject adopted the latter model until about ten years ago. Or, again, although Robbins pointed out, in his [1952], the fascinating possibilities opened up by considering each Neyman-Pearson hypothesis testing problem as one of a group of such problems—so that one had a family of sample spaces, S_i, a family of parameters θ_i, and a family of probability functions p_i, with the suffix i running, say, from 1 to 50—it took more than ten years, and an article by Neyman himself to call the attention of the statistical world at large to the breakthrough (as Neyman described it) which Robbins had achieved. As to the fiducial argument there are, of course, genuine problems associated with defining its domain of validity, but to this day it is remarkable how many writers on the subject try to fit it into either the Bayesian or the Neyman-Pearson model, with which it manifestly is not consistent.

The faithfulness with which mathematicians adhere to a single model in spite of all the difficulties to which such an attitude gives rise tempts one to pun on the word and to call for more permissiveness. A philosophical discussion of fundamentals such as Hacking's book provides will, at the very least, inspire a more reflective point of view.

With the exception of the fiducial argument it is a common property of all the models for mathematical statistics referred to above, that they require no discussion of the semantics of the word 'probability'. For Neyman-Pearson-Wald, for the Bayesian model, and for the compound decision model we merely need to recognise in the situation to which the model is being applied some entity to which the Kolmogorov axioms apply; the nature of this entity may vary, indeed, from instance to instance—it may be a subjective state of mind in one case, an objective state of knowledge in another, or a real or hypothetical frequency in others. In many cases there may coexist a multiplicity of valid interpretations. The fact that statisticians who disagree widely in their attitudes to the foundations of their subject nonetheless tend to agree in their practical advice to scientists and engineers suggests that such ambiguous cases form the majority of those which arise in practice. Such systematic ambiguity—to use Russell's phrase, though semi-systematic ambiguity would here be better—is, of course, characteristic of mathematics.

That Hacking is concerned to discuss at a deeper level, is made clear early in his argument where he says he is concerned to study a property of a ***chance set-up*** on which ***trials*** are conducted, called by some 'long run frequency', or 'chance', and often called 'probability'. A full account of possible meanings of 'probability' is ***not*** attempted. Hacking's 'chance' thus corresponds closely to what Jeffreys calls 'chance' as distinct from 'probability'. It is also closely related to my own notion of 'forward log-odds', or 'f-lods'.[1]

The differences between Hacking's 'chance' and my 'f-lods' arises from two causes (apart from the trivial logarithmic transformations). In the first place, I was concerned to ***derive*** the usual laws of addition and multiplication from more elementary notions; and in the second place I was concerned to preserve logical distinctions which Hacking does not bother with—in particular the distinction between an observable proposition and an hypothesis concerning chance distributions. Thus I borrowed Whitehead's term 'concrescence' (changing its meaning) to denote the conjunction of an observable proposition a and a statistical hypothesis α, to bring out the fact that my fields of observable propositions were closed under conjunction ($a \cdot b$ denoting the observable proposition 'a occurring in the first of a pair of trials and b occurring in the second'—so that $a \cdot b$ should be read as 'a and then b'), whereas putting together an observable proposition and a statistical hypothesis would not produce an observable proposition, but an entity of a different logical kind.[2] Another difference between my own account and Hacking's arises from my own leanings towards constructivism in relation to the foundations of mathematics, which lead me to avoid the use of negation wherever possible, and especially to avoid using the law of excluded middle. I am afraid all these rather irrelevant scruples made the main argument of my [1949] difficult to follow. Since it will be relevant to the discussion of Hacking's views on the fiducial argument, perhaps I may be permitted to outline my own development again here.

I start from the notion of a ***simple proposition***, denoted by a symbol like $a, b, c, \ldots$, which is supposed to represent the ***complete*** description of the result of a trial. Trials are repeatable, and the result of a pair of trials, for instance, could be $a \,.\, b$ (read as 'a and then b'); such a pair of trials can be considered as again a (compound) trial. A statistical hypothesis about trials of a given kind then is taken to establish an ordering of simple propositions as to their plausibility: this ordering is supposed to satisfy certain simple axioms. It is then shown that it is possible to attach a number $l(a)$ to any simple proposition in such a way that $l(a) < l(b)$ if and only if a is less plausible than b.[3] This number obeys the multiplication rule, $l(a \,.\, b) = l(a) \,.\, l(b)$.

The field of simple propositions is not closed under disjunction—to say that a trial resulted in 'a or b' is not to describe the result completely. However, the composite proposition 'a or b' is still observable, and the question arises whether the quantification of plausibility can be extended to cover disjunctions. It was shown later that this can indeed be done, and in essentially only one way. The resulting measure of plausibility then need not satisfy the addition rule, but it must (on certain 'smoothness' assumptions) satisfy a law of the form, for mutually exclusive a and b,

---

[1] Expounded in Barnard [1949].

[2] The set theoretical correlative of the conjunction $a.b$ is the cartesian product, not the intersection of sets a and b. Concrescence has no set-theoretical correlative.

[3] On the way I assumed an Archimedean axiom. I noted at the time that this axiom need not be assumed, and that to omit it would, while complicating the mathematics, enable the theory to cover some interesting possibilities. Such possibilities later received attention from Rényi, in connection with his theory of conditional probability spaces. (Cf. Rényi [1955].).

$$l(a \text{ or } b) = (l(a))P+(l(b))P^{1/p}$$

for some positive value of p. Choosing $p = 1$ can then be justified as being needed to keep the mathematical formulae simple—any other choice would lead to an exactly equivalent theory, looking mathematically more complicated.

The point of this development was twofold. The first aspect, which does not directly concern the discussion of Hacking's work, was to show how a quantitative theory of probability could be derived from weak assumptions about plausibility orderings and the notion of repeatable trials; the second aspect, relevant here, was to show the extent of the duality between the notion of plausibility for observable events (corresponding to Hacking's 'chance') and the notion of plausibility for statistical hypotheses—corresponding to Hacking's 'support'. In fact the development of the theory of plausibility for simple statistical hypotheses was exactly dual to that for simple observable propositions; the parallelism fails just at the transition from simple hypotheses to composite hypotheses. A simple hypothesis is a complete specification of the relative plausibilities of observable results of trials of a given kind; a composite hypothesis is just a disjunction of simple hypotheses. But the difference of logical status between a simple statistical hypothesis and a composite one seems much greater than that between a simple observable proposition and a disjunction of these. For we can always imagine a modification of the given experimental set-up which makes a disjunction of simple propositions itself a simple proposition relative to the modified set-up. In the coin tossing example discussed above we could imagine ourselves not being able to see the results of individual throws, but only the final score recorded on a counter which adds 1 for each H. The new sample space would then consist of the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and the result $r = 1$ would correspond to the disjunction of the 10 alternatives for the original kind of trial such as *HTTTTTTTT*, *THTTTTTTT*, ... *etc.* With the modified set-up, '$r = 1$' would be a simple proposition.

We cannot indicate a corresponding general method for modifying the set-up to convert a composite hypothesis into a simple one. Such modifications are sometimes possible. For instance, the composite hypothesis which consists in saying that a pair (x, y) of measurements is normally distributed with correlation coefficient 0·5 (without specification of the means and variances) can be regarded as a simple hypothesis about the distribution of the estimated correlation coefficient r, calculated by the usual formula from a set of observations, together with a simple hypothesis about the sample configuration. This reduction is possible because the composite hypothesis here considered is invariant under changes of location and scale for x and changes of location and scale for y; we can imagine a change in the experimental set-up under which we become ignorant of the units and origin of measurement of x and of y—for example we could be given a set of pairs such as $(2, -3)$, $(1, 4)$, $(9, 11)$, ... *etc.* in which the first figure is known to refer to the height of an individual, and the second to his weight. The fact that we may not be told from what origin the height measurements are taken, nor whether they are in inches or in centimetres, does not in any way prevent us from assessing the plausibility of their being normally distributed, nor the plausibility of their correlation having any particular value. But in the absence of group invariance, or of some other special structural property having like effect, the reduction of a composite hypothesis to a simple one, by changing the experimental set-up, is not in general possible.

It follows that the quantitative measure of plausibility for simple hypotheses, unlike that for simple propositions, cannot be extended to cover disjunctions, so that while the multiplication rule operates, the addition rule does not. To quote Fisher: 'Whereas such a phrase as "the probability of A or B" has a simple meaning, where A and B are mutually exclusive possibilities, the phrase "the likelihood of A or B" is more parallel with "the income of Peter or Paul"—you cannot know what it is until you know which is meant'. In fact, the dual measures of plausibility, for observable propositions and for simple statistical hypotheses, correspond closely to the Fisherian concepts of probability and likelihood, respectively. But whereas for Fisher the likelihood is, by definition, proportional to the probability, in the theory as developed in my [1949] it was merely assumed that the ratio of relative plausibilities of two simple hypotheses on two different simple propositions was some function of the relative plausibilities of the two propositions on the two hypotheses. The form of the functional relationship was then deduced, and it turned out that the measure of plausibility of a hypothesis, on observation a, would have to equal the plausibility of a, on hypothesis α, multiplied by some function of α. This latter function could be regarded as the 'prior' likelihood of α, to use Edwards's term, though this possibility was not noted at the time. It could also refer to the 'acceptability' of α.

To return now to Hacking, he takes the laws of addition and multiplication for chances as given, and a specification of the numerical chances of possible outcomes of a trial of a given kind is a statistical hypothesis. He then introduces the idea of *support* for a proposition by data, and says (p. 28): 'If one hypothesis is better supported than another, it would usually be, I believe, right to call it the more reasonable. But of course it need not be reasonable positively to believe the best supported hypothesis, nor the most reasonable one.' He goes on to make clear that support is a concept independent of utility. This is, of course, unexceptionable; but the first sentence of the two quoted is open to objection. In a kind of trial in which we expect simple laws to operate we may have a hypothesis involving some weird mathematical function better supported than, say, a simple linear law of relationship. In such a case we might well say that the simple linear law was more reasonable, although the weird law was better supported. Such a discrepancy between support and reasonableness could be accounted for on the theory of my [1949] by the factor referred to in connection with 'prior likelihood', or 'acceptability'; there does not appear to be any corresponding possibility in Hacking's theory. To this extent Hacking's theory appears to be more strongly empirical, in that hypotheses are to be judged purely on the basis of observable data, not on any structural properties they may possess.

At this stage in his argument Hacking inserts a discussion of Wald's decision-theoretic approach to inferential problems. This section is beautifully argued and expressed and should be read with care by any who may still believe that decision theory provides all the answers to problems of inference, as also should be the following chapter, in which 'long run' justifications for rules of inference are demolished. This then leads to the exposition of the 'Law of Likelihood'. Here it is important to note that Hacking's use of likelihood differs from Fisher's, and, to that extent, it fails to correspond to the 'b-lods' notion expounded in my [1949] quite apart from the possibility, already noted, of an 'acceptability factor'. For Hacking, likelihood is a predicate of an ordered pair of propositions, namely a statistical hypothesis and an outcome of a trial; the likelihood is the chance of the outcome if the hypothesis is true. He correctly says (p. 57) that 'Fisher generally uses likelihood as a predicate of hypotheses, in the light of data', but goes on to suggest that Fisher sometimes uses the term in the way Hacking does; I cannot find that this is so.

It should be added that in saying that it is important to note this point, I do not wish to imply that Hacking's argument is in any way weak—apart from the tacit commitment to empiricism already mentioned. The importance derives from the fact that statisticians who are accustomed to Fisher's usage may find it difficult to follow Hacking's reasoning.

Hacking's law of likelihood for discrete distributions involves the notion of a simple joint proposition, which is a proposition of the form 'the distribution of chances on trials of kind K on set-up X is D; outcome E occurs on trial T of kind K'. The likelihood of such a simple joint proposition is what the probability of E would be on trials of kind K if the distribution were D. The law of likelihood also involves the idea of a (composite) joint proposition, which differs from a simple joint proposition in that a class of distributions may be involved instead of only one, and that the kind of trial K' on which E occurs need not be that to which the distributions refer. The statement of the law of likelihood for discrete distributions then is: if h and i are simple joint propositions and e is a joint proposition, and e includes both h and i, then e supports h better than i if the likelihood of h exceeds that of i.

When h and i have the same distributional part, this law says that the occurrence of event E is better supported than the occurrence of F if the probability of E is greater than that of F; when h and i have the same observational part, the law says that the hypothesis H is better supported than the hypothesis H' if the likelihood (in Fisher's sense) of H is greater than the likelihood of H'. Apart from possible considerations of 'prior' 'likelihood', or 'acceptability', these principles seem unexceptionable.

Although Hacking shows that he has read the works of Fisher and of other statisticians with more care than have most statisticians, he fails to do justice to Fisher's conception of likelihood in suggesting that Fisher saw it only as a quantity to be maximised in connection with statistical estimation. The following quotation, from Fisher's introduction to the first edition of his [1925] clarifies the situation: 'This is not to say that we cannot draw, from knowledge of a sample, inferences respecting the population from which the sample was drawn, but that the mathematical concept of probability is inadequate to express our mental confidence or diffidence in making such inference, and that the mathematical quantity which appears to be appropriate for measuring our order of preference among different possible populations does not in fact obey the laws of probability. To distinguish it from probability, I have used the term "Likelihood" to designate this quantity, since both the words "likelihood" and "probability" are loosely used in common speech to cover both kinds of relationship.'

The next chapter of Hacking's book is concerned with 'tests of significance'. Here there seems to be a major flaw. He quotes, approvingly, W. S. Gossett as saying that a test 'doesn't in itself necessarily prove that the sample is not drawn randomly from the population even if the chance is very small, say ·00001: what it does is to show that if there is any alternative hypothesis which will explain the occurrence of the sample with a more reasonable probability, say ·05 (such as that it belongs to a different population or that the sample wasn't random or whatever will do the trick) you will be very much more inclined to consider that the original hypothesis is not true'. He takes this passage by way of rebuttal of Fisher, whom he accuses of personal criticism of those who stressed the importance of alternatives to the hypothesis being tested, in connection with the theory of testing. In fact Fisher referred approvingly to the concept of the power curve of a test procedure and although he wrote: 'On the whole the ideas (*a*) that a test of significance must be regarded as one of a series of similar tests applied to a succession of similar bodies of data, and (*b*) that the purpose of the test is to discriminate or "decide" between two or more hypotheses, have greatly obscured their understanding', he was careful to go on and add 'when taken not as contingent possibilities but as elements essential to their logic'. It would take too long to give a full discussion of the logic of significance tests here; a reference to Anscombe's paper (Anscombe [1962]) must suffice. The difficulty with Hacking's account is that it leads him to say (p. 89): 'An hypothesis should be rejected if and only if there is some rival hypothesis much better supported than it is.' But there *always* is such a rival hypothesis, *viz.* that things just had to turn out the way they actually did.

The fact seems to be, that the domain of application of Hacking's theory, like that of the theory of likelihood, is confined to comparisons between statistical hypotheses. Situations in which only one hypothesis is in question, and the issue is whether agreement between this hypothesis and the data is so bad that we must begin to exercise our imaginations to try to think of an alternative that fits the facts better—tests of goodness of fit, in other words—are not satisfactorily dealt with by this theory.

Hacking's next chapter discusses the Neyman-Pearson approach to hypothesis testing and he reaches the conclusion, with which I agree, that this approach is best regarded as a theory applicable to the advance planning of experiments. On a purely personal point, I feel bound to comment on what is said on page 103: 'The idea that Neyman-Pearson tests could better serve before-trial betting than after-trial evaluation has lain fallow in a remark made by Jeffreys twenty years ago. Last year [*i.e.* in 1963] Dempster produced, for the first time in all that period, a similar distinction'. Hacking here seems to have overlooked my [1950] from which I quote: 'it is the view of the present writer that the arbitrary nature of the reference set involved, on the Neyman-Pearson theory, in a test of significance, is a decisive reason for rejecting that theory, as a theory of inference, in favour of using a theory of inference, such as that given by Fisher, where the idea of a reference set does not enter. It should be emphasised, however, that the Neyman-Pearson theory is an exceedingly valuable weapon in the advance planning of experimentation. To put the matter shortly, two kinds of quantity are involved in uncertain inference. The present writer has called them f-odds and b-odds, but Fisher has more aptly (though slightly less precisely) called them probability and likelihood. Probabilities are relevant before an experiment has been performed, when we are planning it. After the experiment has been performed, when we are drawing conclusions, likelihoods are relevant. As a theory based on probabilities, the Neyman-Pearson theory is useful in planning, before the result is known; but after the result is known the theory of likelihood should be used.' Although I was conscious, when I wrote about likelihood in my [1947] that its use would be restricted to the comparison of one hypothesis with another, and was at that time and again in my [1949] careful to say so, in the passage quoted above I incautiously omitted to refer to this.[1]

Hacking's next chapter discusses randomness, in connection with random sequences and with random sampling. He refers to Church's suggestion to overcome the difficulties in von Mises's idea of a random sequence by using the fact that the set of computable functions is countable; but he does not appear to have noticed Ville's destructive criticism of this idea. Ville's point essentially was that any attempt to define mathematically the class of random sequences of 0's and 1's was doomed to failure, because, representing such sequences as binary fractions, one would be excluding as non-random any set of such fractions of Lebesgue measure zero. But however many such sets had been removed from the unit interval, the remaining set would, of course, have measure 1, and so would contain infinitely many more sets also of measure zero, and these sets ought also to be excluded. This argument persuaded many, including myself, that the search for a mathematical definition of randomness was doomed to failure, until Martin-Löf came along with an application of the idea of computability which went in a somewhat different direction from Church. Briefly, Martin-Löf's definition of a (finite) random sequence is, that it is a sequence which is computed by a Turing machine whose Gödel number (in the binary scale) is as long as the sequence itself. He overcomes Ville's difficulty by pointing to the fact that the set of *computable* sets of measure zero is countable, so that the union of all such sets also has measure zero. It is this union of all such sets which is removed from the unit interval.

Apart from brief references to random sequences, Hacking's main concern in this chapter is to discuss random sampling from finite populations in a way which does not seem to differ markedly from standard statistical doctrine. It is the next chapter, entitled 'The fiducial argument' which has perhaps drawn forth the most comment from reviewers, in that Hacking sets out a mode of reasoning leading effectively to probability statements about hypotheses (though Hacking sticks to the term 'support', rather than probability).

He draws a fruitful analogy with Frege's foundations for set theory and arithmetic, in which Russell found a contradiction. He says (p. 151) 'Just as (after Russell's criticism of Frege) we needed a more adequate characterisation of irrelevance. Each genius (Fisher and Frege) proferred a false conjecture. The task of future generations is not to confute the genius but to perfect the conjecture. Our principle of irrelevance is some way along that road.' Indeed it is, as also is the work of Sprott, and of Fraser, to which Hacking does not refer. But in his extended list of contents, Hacking says 'It is proved that any principle similar to the principle of irrelevance but stronger than it, must lead to contradiction. So the principle completes a theory of support.' In the chapter in question, however, Hacking makes this idea plausible, but could not be said to have produced a proof. My own conjecture here is, that the situation will continue to be analogous to that holding in the foundations of set theory, where, in a sense, the search for a definitive formulation has definitively been shown to be fruitless. One is, in fact, tempted also to conjecture that, just as Hadamard said that his brain was so constructed that the axiom of choice appeared obvious to him, while to Borel it was far from obvious, and Cohen has shown that adding either the axiom of choice or its negation to the other axioms of set theory will not produce a contradiction, so we may find, when the rules of statistical inference are more fully formalised than now, that options remain.[1] Such a conjecture looks more plausible when we bear in mind that the appeal, for example in Hacking's theory, to the notion of a trial *of kind K* involves an act of class formation.

---

[1] Also, cf. Barnard [1951] and Barnard, Jenkins and Winsten [1962].

Hacking's next two chapters are concerned with estimation. He points to difficulties in the standard theories, but does not always go as far as he might have done. For example, on page 183 he says: 'Of course it might be true that the best of unbiased estimators were better than estimators of other kinds, even though inferior unbiased estimators are inferior to other estimators. But this has never been proved.' Indeed, there are counter examples. In inverse sampling to estimate a single probability θ it can be shown that the *only* unbiased estimator is restricted to taking the values 0 or 1; but if the sample space is the open unit interval, or, for instance, restricts θ to lie between $\frac{1}{3}$ and $\frac{2}{3}$, this means that the only values the unbiased estimate can take are known to be impossible ones. Hacking makes an interesting suggestion, that in asking for an estimate we should always specify a scale of measurement, and this leads him to reject the idea that in estimating θ we are necessarily at the same time estimating any function $f(\theta)$, and that our estimate of $f(\theta)$ must be $f(t)$ if our estimate of θ is t. This amounts to saying, in the Neyman-Pearson model, that the parameter space should be endowed with a metric, rather than being left as an unstructured set. Unfortunately Hacking does not develop this idea to any extent. My own guess would be that we should think of the rather weaker notion of a topology on the parameter space (a metric implies a topology, but not conversely), so that we can give a meaning to the idea of converging on the true value of θ when the sample size becomes larger and larger; in some cases a family of locally equivalent metrics—intermediate between a single metric and a topology—seems appropriate. Le Cam has used ideas of this kind, in connection with what Neyman termed 'Best Asymptotically Normal' estimates; his theory can be regarded as a further development and formalisation of the earlier part of Fisher's [1925*a*].

It is a pity that Hacking does not discuss the idea, eminently practicable now that we have computers with graphical output facilities, of regarding the graph of the likelihood function (or, equivalently, the graph of its logarithmic derivative) as answering the problem of estimation. It will indicate the value or values to which the data point most strongly (*i.e.* which are best supported, or most likely), it will indicate with what precision the data point to the neighbourhood of a particular value, and it will preserve all the relevant information in the data, in the sense that it is a minimal sufficient statistic.

The final two chapters are devoted to discussion of Bayes's theories, first in their 'objective' versions as represented by Bayes and by Jeffreys, then in the subjective versions associated with de Finetti, Ramsey and Savage. The objective versions are rejected 'for reasons which by now are entirely standard', while the subjective theory is explained 'sympathetically'. In so far as Hacking's theory is restricted to cases in which the notion of 'chance' is involved, its domain is narrower than that of the subjective Bayesians; at the same time it is more explicit in its application to problems of inference in the natural sciences.

---

[1] The possibility that *valid* theories of statistical inference could be genetically determined is worth meditating upon.

The last chapter shows signs of having been written some time after the earlier ones, and it seems to shift the emphasis in places. For example, on page 222 Hacking comes near to discussing a 'goodness of fit' situation, and says 'My theory of statistical support does not attempt rigorous analysis of the reasoning here.... The theory of statistical support cannot judge the force with which an experiment counts against a simplifying assumption.' To this extent he appears to agree with the comment made above on his treatment of tests of significance. Again, on page 219, Hacking appears to entertain the possibility of something corresponding to the idea of 'prior likelihood' of 'acceptability' referred to above, and he explicitly refers to the point made by Fraser (and also by the present writer) that group invariance and other structural features of an experimental set-up may be relevant to its statistical interpretation.

It is clear that in this area there is much further exploration to be done. Hacking's book remains an invaluable guide book for anyone willing to join in this task.

G. A. BARNARD
University of Essex

REFERENCES

Anscombe, F. J. [1962]: 'Tests of Goodness of Fit', *Journal of the Royal Statistical Society* (*B*), **25**, pp. 81–94.

Barnard, G. A. [1947]: Review of Abraham Wald: 'Sequential Analysis', *Journal of the American Statistical Association*, **42**, pp. 658–64.

Barnard, G. A. [1949]: 'Statistical Inference', *Journal of the Royal Statistical Society* (*B*), **11**, pp. 115–49.

Barnard, G. A. [1950]: 'On the Fisher–Behrens Test', *Biometrika*, **37**, pp. 203–7.

Barnard, G. A. [1951]: 'The Theory of Information', *Journal of the Royal Statistical Society* (*B*), **13**, pp. 46–64.

Barnard, G. A., Jenkins, G. M. and Winsten, C. B. [1962]: 'Likelihood Inference and Time Series', *Journal of the Royal Statistical Society* (*A*), **125**, pp. 321–72.

Fisher, R. A. [1925*a*]: 'Theory of Statistical Estimation', *Proceedings of the Cambridge Philosophical Society*, **22**, pp. 700–25.

Fisher, R. A. [1925*b*]: *Statistical Methods for Research Workers.*

Rényi, A. [1955]: 'On a New Axiomatic Theory of Probability', *Acta Mathematica Academiae Scientiarum Hungaricae*, **6**, pp. 285–335.

Robbins, H. [1952]: 'Asymptotically Sub-Minimax Solutions of the Compound Decision Problem' in J. Neyman (*ed.*): *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 131–48.

LIKELIHOOD

The fundamental question about statistical inference is philosophical: what primitive concepts are to be used? Only two answers are popular today. Edwards is the first scientist to write a systematic monograph advocating a third answer.[1]

---

[1] Edwards, A. F. [1972]: *Likelihood. An Account of the Statistical Concept of Likelihood and its Application to Scientific Inference.* Cambridge: Cambridge University Press. £3.80. Pp. xiii+235.

'Orthodox' statistics admits only one primitive concept, namely physical probability, a propensity that betrays itself in stable long run relative frequencies. The statistician devises procedures for testing hypotheses and estimating parameters; he chooses among possible procedures by pointing out desirable properties that show up under repeated sampling. There is much infighting about what properties *are* desirable—unbiasedness, minimum variance, size, power, significance level and all that—but all these properties have to do with long run operating characteristics of statistical procedures. The orthodox statistician, be he a follower of Neyman or Fisher or whoever, will not usually allow you to measure the credibility of any particular estimate or hypothesis. He allows you to say only that this particular estimate was made by a procedure that has virtues that would show up under repeated sampling, or, for example, that this hypothesis is rejected by a procedure that mistakenly rejects hypotheses only one per cent of the time.

Bayesian statistics tells quite the opposite story. The only primitive concept is degree of belief, which arguably ought to satisfy the probability calculus. If you do have probabilistic degrees of belief in the propositions in some field of interest, there is a simple model of learning from experience that Bayesians find satisfying. Orthodox statistics is characterised by a host of locally applicable, often *ad hoc*, but notably ingenious procedures. A single, simple global analysis is the delight of the Bayesian.

Thomas Bayes's original paper was published in 1763. The use of repeated sampling properties in inference goes back to Jacques Bernoulli, published 1713. Between these dates J. H. Lambert may have toyed with a third basic concept, although he subsequently developed the theory of errors in what I would now call the 'orthodox' way. In 1777 Daniel Bernoulli brought this third alternative more into the open. To take his engaging example, we know that one of two archers has been firing at a target. One is an untalented novice, the other a master. We observe that the shots cluster around the bull. Who aimed? Knowing the abilities of the two men, we reason: if the novice fired, then something very unlikely must have happened, but if it was the master, an event of far greater probability has occurred. On this evidence, we strongly favour the hypothesis, that the master shot the arrows at this target. If we restrict 'probability' to the orthodox, physical, usage, we cannot say anything about the *probability* that this is the target of the master, not the novice. (We could if we knew that the two men tossed coins to decide who would shoot, but that is not part of our data.) We can, however, follow D. Bernoulli and compare what R. A. Fisher called the *likelihoods* of the two hypotheses. Likelihood is a sort of inverse of physical probability. The likelihood of the hypotheses h, in the light of data e, is the probability of observing e if h were in fact true. D. Bernoulli apparently advises us to prefer hypotheses of greater likelihood given the data.

Euler at once retorted that this advice is metaphysical, not mathematical. Quite so! The choice of primitive concepts for inference *is* a matter of 'metaphysics'. The orthodox statistician has made one metaphysical choice and the Bayesian another. D. Bernoulli appears to have been proposing a third. Likelihood is of course formally defined in terms of probability, but it is being offered as a primitive concept of inference; primitive in the sense that it is supposed to justify inference, and that its use is supposed to need no further justification.

As a primitive concept likelihood did not fare very well, although one finds it suggested in surprising quarters. Edwards reminds us that John Venn, a good frequentist, seems to be prepared to use likelihood as a primitive tool in assessing statistical hypotheses, and that F. P. Ramsey, who put the subjective theory on a solid modern footing, also, in a late paper, comes down in favour of likelihood. But for all these hints the idea of likelihood would have lain fallow had it not been for Fisher. And Fisher was completely ambivalent about the concept. In informative asides Edwards reminds us of this ambivalence and in an historical paper (submitted to *Biometrika*) he has expanded this theme into a thorough history of likelihood.

It is important not to confuse Fisher's 'method of maximum likelihood' with likelihood as a primitive concept. The former is a valued technique in estimation theory, and can be justified by a large number of long run sampling properties. Fisher himself discovered many of these properties, and made the method of maximum likelihood an integral part of orthodox statistics. Equally, of course, likelihood is a crucial quantity for the Bayesian, who learns from experience by multiplying prior probabilities and likelihoods. But for neither orthodox nor Bayesian statistician is likelihood primitive; it is only a quantity that turns out to be central in many calculations. Fisher, from time to time, wanted likelihood to do more.

Fisher had no use for Bayesian analysis unless one is confronted by that rare situation in which hypotheses of interest are themselves generated randomly from a chance set-up. So he could not, in general, speak of the probability of an hypothesis. But he urged, from time to time, that relative likelihoods of hypotheses form the natural way to indicate the relation between hypotheses and evidence. This idea is present in some of the great work of the early 1920s, and recurs somewhat nervously in the middle thirties; it comes out again in *Statistical Methods and Scientific Inference*, his final major contribution. Fisher even goes so far as to say that likelihood is much like Keynes's idea of logical probability, except that it is not subject to the unjustifiable addition law for probabilities: we cannot add likelihoods to get the likelihood of some disjunction of hypotheses. Harold Jeffreys once told Fisher there was nothing wrong with postulating likelihood as a basic axiom for inference, and as Edwards puts it, Fisher 'wistfully' contemplated that, but was never altogether sure it was the right thing to do. Edwards, in contrast, has no doubts.

Edwards's book is more of a sermon than an attempt to provide logical foundations for a new mode of argument. Himself a geneticist, he is appealing to fellow scientists to reason in a certain way. He gives plenty of attractive examples, for the theory is to be tested by having good consequences for science. Edwards clearly has a strong 'intuition' that likelihood is the right tool, but unlike philosophers who boringly go on about their intuitions, Edwards aims at establishing a coherent body of method that enables the scientist to analyse his data in a sensible way.

Concerning the philosophical question, 'What primitive concepts to use?', Edwards is at first orthodox. There is only one kind of probability, the kind that shows up as stable relative frequency. He grants, fleetingly, that Bayesians have arguments showing that if one has degrees of belief in every sort of proposition, or if one is made to bet on anything under the sun, then in coherence one's betting rates will be probabilities. But he retorts that he simply does not have degrees of belief in the genetical hypotheses he contemplates, and no one forces him to lay bets on which hypotheses are right. Indeed the statistical models used in genetics are not properly called true or false; they are more or less adequate and there is no sense in betting on their truth.

Edwards tells his colleagues not to bet and so not to be Bayesian. He asks them to assess the relation between experimental evidence and statistical models of interest. He will not accept procedures justified merely by their long run operating characteristics: he wants to know how *this* particular piece of evidence bears on *these* hypotheses offered by *this* working model. Scientific inference, he believes, has essentially to do with particular cases. So orthodox statistical procedures are to be rejected. Edwards is left with likelihood.

He finds it convenient to measure relative likelihoods of hypotheses by taking natural logarithms, and he calls (relative) log-likelihood the measure of (relative) support for hypotheses by data. So to compare the support that e furnishes for h_1 against h_2, compare the logarithm of the probability of getting e, according to h_1, with the logarithm of the probability of getting e, according to h_2. Logarithms are used because this makes support 'additive' in a certain sense. If we have two independent pieces of data bearing on some hypotheses, the log-likelihood of the two items taken together is the sum of the individual log-likelihoods. So although we cannot combine the support for different hypotheses, we can combine different pieces of support for the same hypothesis. This, says Edwards, is exactly what we want in science, for a disjunction of hypotheses is no hypothesis at all. I may contemplate the hypothesis that the refractive index of a crystal is r, and the hypothesis that it is r', but there is no scientific hypotheses to the effect that the refractive index is r-or-r'. Philosophers, however, know to their cost that there is no good way to distinguish 'real' hypotheses from 'manufactured' ones. One can think of real cases as well as logicians' tricks. Surely I can contemplate the hypothesis that a certain quantity has a distribution from a specified part of a family of distributions (normal or log-normal or whatever) without having much idea about, or even interest in, specific parameter points. Can I not then ask how well supported this 'composite' hypothesis is?

In the case of simple hypotheses there is still a question about the actual log-likelihood numbers. Do they mean anything? The orthodox statistician will say he does not know what to *do* with them. Probabilities and operating characteristics let you do all sorts of things. Given some utilities, they allow you to compute expected loss. But what can we do with likelihoods? Edwards's reply is twofold. First, he does not want to do any of the things orthodox statisticians can do. He attaches no sense to a loss function over hypotheses in genetics. He wants to report evidence and show, in brief form, how it bears on the hypothesis under examination. All right, but what do Edwards's numbers mean? He has an unexpected but sensible answer. Use likelihoods and you will find out from experience what the numbers mean. It takes a while to learn, for example, what temperatures mean, and it is notoriously hard for Fahrenheiters to take in weather forecasts in Centigrade, but spend a summer in the South of France and you get to know what 30°C feels like. Numbers that basically record a ranking take a lot of getting used to. Indeed we could make Edwards's remark even about probabilities. Shortly before his death L. J. Savage was saying that we have only just begun to get a grasp of personal probabilities, and it might take several generations more before they were properly entrenched in our understanding. Whether or not he is right, it is clear that numerical probabilities have been radically extending their domain and their intelligibility. A meteorological classic from the early decades of this century tells the forecaster never to use the word 'probable' for it would be redundant in a weather forecast, all forecasts being merely probable. Nowadays the U.S. Public Weather Office seems incapable of uttering a sentence not qualified by a probability number. Savage, incidentally, thought the Weather Office simply did not know what its own probability percentages meant. Edwards could carry his fight into the enemy camp: if you use likelihoods in reporting experimental work, you will get to understand them just as well as you now think you understand probabilities.

Many philosophers will resent this kind of reasoning but I do not find it intrinsically disturbing. What does worry me is Edwards's faith that a given log-likelihood ratio will mean the same in any circumstance whatsoever. Grant, for a moment, that if the likelihood of h_1 on e exceeds that of h_2, then, lacking other information, e supports h_1 better than h_2. Now suppose the actual log-likelihood ratio between the two hypotheses is r, and suppose this is also the ratio between two other hypotheses, in a quite different model, with some evidence altogether unrelated to e. I know of no compelling argument that the ratio r 'means the same' in these two contexts. Physical probability is much better off. There is always the frequency interpretation telling us that if the probabilities of two unrelated events, in different chance set-ups, are the same, then the two events tend to occur equally often. That, I think, is the chief virtue of the frequency interpretation: it shows that different probabilities in different set-ups are commensurable. No non-Bayesian argument shows that likelihood ratios in different situations are always commensurable, that is, measure the same levels of evidential significance.

Indeed in artificial cases there seem to be positive counterexamples to unrestrained use of likelihood. A classic case is the normal distribution and a single observation. Reluctantly we will grant Edwards that the observation x is the best supported estimate of the unknown mean. But the hypothesis about the variance, with highest likelihood, is the assumption that there is *no* variance, which strikes us as monstrous. Edwards is a practical reasoner, and is inclined to disregard this case. If we do wish to fit it into the likelihood scheme of things, we must concede that as prior information we take for granted that the variance is at least w. But even this will not do, for the best supported view on the variance is then that it is exactly w.

For a less artificial example, take the 'tram-car' or 'tank' problem. We capture enemy tanks at random and note the serial numbers on their engines. We know the serial numbers start at 0001. We capture a tank number 2176. How many tanks did the enemy make? On the likelihood analysis, the best supported guess is: 2176. Now one can defend this remarkable result by saying that it does not follow that we should estimate the actual number as 2176, only that comparing individual numbers, 2176 is better supported than any larger figure. My worry is deeper. Let us compare the relative likelihood of the two hypotheses, 2176 and 3000. Now pass to a situation where we are measuring, say, widths of a grating, in which error has a normal distribution with known variance; we can devise data and a pair of hypotheses about the mean which will have the same log-likelihood ratio. I have no inclination to say that the relative support in the tank case is 'exactly the same as' that in the normal distribution case, even though the likelihood ratios are the same. Hence even on those increasingly rare days when I will rank hypotheses in order of their likelihoods, I cannot take the actual log-likelihood number as an objective measure of anything.

Edwards's book has many technical and practical suggestions beyond the scope of this journal. There is one 'philosophical' novelty to be remarked: the device of imaginary experiments. This has had some small play in Bayesian literature, and perhaps goes back to Laplace, but it is newly introduced into likelihood. Suppose I am thinking about some hypotheses and have not, as yet, got any new data. Still, I am not indifferent between the hypotheses; I have background information and prejudices. I realise, perhaps, that I regard h_1, in comparison to h_2, *as if* I had evidence e conferring a log-likelihood ratio r between these two hypotheses. This, then, is my 'prior support'. When I do get some real data, I can add the prior support and the actual experimental support to get a posterior support.

Edwards's very definition of prior support reads oddly. 'The *prior support* for one hypothesis against another is S if, prior to any experiment, I support the one against the other as if I had conducted an experiment...' (p. 36). In most of the book it is *data* that do the supporting, but here it is *me!* After I have done a real experiment, I am supposed to add 'my' prior support to the support furnished by the evidence; that looks dangerously like using arithmetic to add pints of milk to pounds of apples. Edwards notes that typically prior support counts little towards posterior support, so this may sound like the Bayesian situation with prior and posterior probability. But the Bayesian prior is the same kind of thing as the posterior, and moreover the prior is needed to get a posterior. Edwards's prior support seems to me a different thing from the support that evidence furnishes for hypotheses, and the latter kind of support does not require the former.

Edwards's book is nicely written, has a straightforward development, plenty of examples and instructive historical asides. It is not a book written for philosophers, but it is a book that those who care about probability ought to read. It gives a puzzling and possibly fundamental inferential concept a longer run than anything published before now. Statistical reasoning is not well enough understood by anyone, yet, and we need more fundamental concepts in the arena than is currently the case. I do not know how Edwards's favoured concept will fare. The only great thinker who tried it out was Fisher, and he was ambivalent. Allan Birnbaum and myself are very favourably reported in this book for things we have said about likelihood, but Birnbaum has given it up and I have become pretty dubious. George Barnard is the only worker who has consistently and persistently advocated and advanced a likelihood philosophy. I hope Edwards's book will encourage others to enter the labyrinth and see where it goes.

IAN HACKING
Cambridge University

THE SUBJECTIVE THEORY OF PROBABILITY[1, *]

1 *Introduction.*
2 *Basic Model.*
3 *Axioms and Coherence.*
4 *Finite versus Countable Additivity.*
5 *Independence and the Laws of Large Numbers.*
6 *Exchangeability: Objective Probabilities in a Subjective Theory.*
7 *Inductive Reasoning.*
8 *Mathematical Aspects.*

1 *Introduction.*

The subjective theory of probability emerged as a solution to certain problems which had arisen in the older logical view of probability. According to this logical account, a body of evidence *e* justifies just one rational degree of belief in a hypothesis or prediction *h*. This is the logical probability of *h* given *e*. Probability is a relation which we can 'cognise correctly' (to use Keynes's phrase ([1920], p. 5)).

The difficulty here is that (as Ramsey puts it ([1926], p. 161)) 'there really do not seem to be any such things as the probability relations he (Keynes) describes'. Admittedly attempts had been made to evaluate the supposed logical probabilities by using the 'principle of indifference'; but apparently insuperable difficulties were found in this principle. Another weakness was that the axioms of probability were justified only by unsatisfactory appeals to 'logical intuition'.

In the face of these problems, the founders of the subjective theory (Ramsey and de Finetti) proposed independently, and at about the same time, that the apparently non-existent objective degrees of rational belief should be replaced by the degrees of actual belief held by particular persons. They proposed an operational method of measuring such degrees of belief, and showed how the axioms of probability could be derived from their point-of-view. They thus established the subjective theory as a viable account of probability.

De Finetti has continued to develop the subjective theory with the help of many distinguished collaborators—notably L. J. Savage. To date however his accounts have taken the form of a number of necessarily fragmentary papers and short monographs. But now he has attempted to draw these fragments together and present something more like a complete development of the subjective view. His book will therefore be welcomed by all serious students of probability theory, and perhaps especially by those more interested in the philosophical aspects of the subject. De Finetti stresses that, in the present volume, he deals mainly with conceptual or foundational questions, and indeed he hopes to produce a second work of about the same length, concentrating on the mathematical side of the

---

[1] Article Review of de Finetti [1970]: *Teoria delle Probabilità, Sintesi Introduttiva con Appendice Critica*, Turin. Pp. 770.

[*] In preparing this review I have received a great number of helpful comments particularly from Jon Dorling, Ian Hacking, Colin Howson, Imre Lakatos, Moshe Machover, John Watkins and Elie Zahar, and especially from Professor de Finetti himself who was kind enough to read the whole of a previous version of the review, and to make very detailed comments which, I hope, have helped to eliminate some earlier misunderstandings of his position.

theory. It is not to be inferred from this, however, that the present volume is devoid of mathematical content. In fact there are many interesting mathematical discussions, some of which are mentioned in section **8** below.

The book has, to date, only appeared in Italian, but an English version will be forthcoming shortly.[1] The main purpose of this review, therefore, is to give an idea of the contents of the book to English speakers. It will be difficult for me, however, to carry out this task in an unbiased fashion, since, as a matter of fact, I adhere to the objective approach to probability developed in the works of von Mises, Popper and others. Rather than concealing these prejudices, it seemed to me better to state them frankly, not only summarising de Finetti's position but also setting down such critical comments as might be made from the opposite, or objective, point of view. To date the statement of de Finetti's views probably best known in the Anglo-Saxon world is contained in his classic [1937], an English translation of which appears in H. E. Kyburg and H. Smokler (*eds.*): *Studies in Subjective Probability*. I shall therefore from time to time compare de Finetti's treatment of a certain question in the present book with his earlier treatment in this paper.

2 *Basic Model.*

In Chapter 2[2] and at the beginning of Chapter 3[2] de Finetti introduces what I shall call his basic model. It is assumed that a certain individual at a given time has a certain amount of knowledge. When he cannot, from this body of knowledge, infer that a particular proposition is true or false, that proposition is possible for him (and he may assign a probability to it). Suppose, for example, that the proposition states that a certain event has occurred. If the individual in question does not know that the event has occurred, it becomes for him a 'random event' (*evento aleatorio*)—though, for another individual, who may perhaps know that the event has occurred, it is no longer a 'random event' but rather a sure event. Thus, as de Finetti himself says, (p. 38)[2]: '. . . the meaning that we give to "random" (*aleatorio*): is simply "unknown" . . .' (to a particular individual). However the event is (p. 38) 'in itself *well defined* (*ben determinato*); . . . to say that it is *well defined* means . . . that it can be specified in such a way that any bet (or insurance) based on it can be settled (at the appropriate moment) in an indisputable fashion.'

Simple though this model is, it is yet liable to objections. It requires that any individual distinguishes between *what he knows*, and what is for him *uncertain but possible*. However it is difficult in practice to make this distinction. Should we, for example, include among what we know generally accepted and well-corroborated laws of modern science? Common sense might well answer: 'yes', and de Finetti seems to agree when he says (p. 37): '. . . your information and knowledge will allow you in general to exclude some of them (*i.e.* of the *a priori* possibilities) as impossible (it will permit you that is—and it has been said that this is the function of science—to set a "limit to what you expect")'. In other words certain *a priori* possibilities can be excluded because they conflict with scientific laws taken as 'known'. On the other hand no scientific law is immune from doubt. It is generally agreed that Newton's theory has been the most successful theory in the history of science. Yet there was scarcely a time when it was wholly free from difficulties and experimental anomalies; and eventually it came to be definitely superseded.

---

[1] The translations of passages from the book in this paper are however my own, and may therefore differ from those in the eventual English version of the full work.

[2] Where chapter and page references are given without further specification, it must be understood that they refer to the book under review.

This line of thought suggests that we should exclude general scientific laws from what we know, and perhaps include only particular results of experiments. Yet even such particular results are not immune from doubt. Experimental results accepted for long periods of time have later been shown to be incorrect through the neglect of some factor. What then *do* we know? Are we to be left with reports of immediate sensations, with '*cogito ergo sum*' or what?

The same argument tells against the introduction of events which are *well-defined* (*ben determinati*) in de Finetti's sense *i.e.* events for which it can be determined in a finite time and 'in an indisputable fashion' whether they occurred or not. It is in fact impossible to establish the truth of any scientific proposition 'in an indisputable fashion'. Even the best-designed experiment can later be shown to have given inaccurate and misleading results. Admittedly in the context of artificial gambling situations such as horse racing we do have 'well-defined' events, because a special committee is set up to decide the matter, and from the verdict of such a committee there is no appeal. No such privileged committee exists for scientific statements which are always open to doubt and criticism.

The points raised so far may seem to be quibbles, but this is not in fact so. The question of whether general scientific laws can be included in 'what we know' is important in assessing the claim of the subjective theory to 'solve the problem of induction' (see Section 7, *below*). While the problem of 'well-defined events' raises the question of whether the subjective approach can be extended from gambling situations such as horse racing to deal with probabilities as they appear within scientific theories (for example in physics). Of course it is the objectivists' view that such an extension cannot be made.

3 *Axioms and Coherence.*

After introducing his basic model, de Finetti now proceeds to probabilities and the establishment of the axioms of probability. In his [1937] the procedure was roughly this. We take one of the well-defined events E, and suppose that for a certain person (A say) it is a random event, *i.e.* A does not know whether E or $\sim E$ is the case. We next suppose that A is forced to bet against an opponent B on E. A has to assign a 'betting quotient' q on E, while B chooses a sum S which may be positive or negative. A has to pay B qS in return for S if E proves to be the case. The betting quotient q is then identified with the subjective probability of E which A assigns relative to his state of knowledge. Such betting quotients are said to be coherent if it is impossible for B to choose the stakes S such that he wins whatever happens. It can then be shown that coherent betting quotients satisfy the axioms of probability (this result is often called the Ramsey–de Finetti theorem). It provides the justification for the axioms of probability in the subjective approach.

The present book introduces a variant on the 1937 treatment. To begin with de Finetti takes as fundamental not the probability $p(E)$ of a random event E, but the mathematical expectation $p(X)$ of a random number (*numero aleatorio*) X.

De Finetti calls a mathematical expectation a *prevision* (*previsione*) for a number of reasons. First of all he wants to make a contrast between the *prevision* of the value of an unknown number and the *prediction* (*predizione*) of that value. Making a prevision he regards as a rational activity: while making a prediction is often mere guessing or pretending to know what we don't know. Bogus predictions he comments (p. 86) are made by 'not only self-styled Magi and prophets, but also experts and other persons inclined to anticipate the future in the forge of their fantasy'. Where we do not know, we should not pretend to knowledge by making a prediction, but rather be rational and make a prevision.

When considering predictions of the results of general elections or of the end of the world, one feels great sympathy with de Finetti's remarks. The difficulty, however, is whether one can clearly demarcate such bogus predictions from for example, a prediction of the moon's position in a month's time. The latter kind of prediction is of course made with the help of well-confirmed theories and elaborate calculations, but the element of guess-work is far from eliminated, and indeed such astronomical predictions have often proved to be false. De Finetti has two possible courses of action. He must either explain how such 'genuine' predictions differ from those which he denounces as mere guesswork, formed 'in the forge of someone's fantasy', or he must claim that even such genuine predictions should be replaced by previsions. But such a suggestion would involve the rewriting of many books of astronomical tables.

A second reason for de Finetti's use of the term 'prevision' is that it allows one to exhibit the analogy between the prevision $p(X)$ of a random number X, and the probability $p(E)$ of a random event E. Indeed let us define the indicator X_E of an event E as the random number which takes the value 1 when E is the case and 0 otherwise. Then if we introduce prevision as the primitive notion, we can define the probability of an event E as the prevision of its indicator X_E. De Finetti regards this approach as simpler and more natural than the standard one of beginning with probabilities and defining previsions by means of integrals.

Starting then with previsions he introduces a new definition of coherence which can be shown to be formally equivalent to his old one. Given a number X which is random for A (the individual assigning subjective previsions), A has to choose a value $\bar{x}$ for X (the prevision of X) on condition that he submits to a penalty proportional to the square of the difference between $\bar{x}$ and X. A set of previsions are then said to be *incoherent* if there is another set which will give a lower loss whatever happens, and *coherent* otherwise. From this definition the axioms of probability (or rather of prevision and hence of probability) can be derived just as before. Although formally equivalent to the old definition of coherence, the new one may have certain advantages from a philosophical point of view. As de Finetti points out (p. 112) it eliminates the need for introducing an opponent, and hence avoids certain problems, *e.g.* the question of whether we should alter our bets in the light of what we know or conjecture about our opponent's opinion or intellect.

Both definitions of coherence raise the question of 'money versus utility'. In what quantities are we going to bet: units of money or units of utility? If we say 'money', we have the usual problem that a £100 bet is less critical for the millionaire than for the ordinary man. The same difficulty arises, though in a less obvious fashion, starting from the second definition. Here, as de Finetti points out (p. 112), we implicitly assume $p(aX) = a\,p(X)$, and this implies the same 'rigidity in the face of risk' which is implausible if we use money to measure value. On the other hand if we make our bets or penalties in terms of 'utilities', we have the problem of measuring utility. De Finetti is sympathetic to the idea of developing a utility theory, but suggests that it is simpler for his own purposes to think of the bets, penalties, *etc.* as being in terms of very small amounts of money.

So far I have loosely said that de Finetti obtains 'the axioms of probability'. However there is a difference between his axiomatization and the standard one. From his axioms it follows that:

(1) If E is a random event $0 \leqslant p(E) \leqslant 1$

(2) If $E_1, \ldots, E_n$ are incompatible events (*i.e.* such that no two can be simultaneously the case), then

$$p(E_1 \cup E_2 \cup \ldots \cup E_n) = \sum_{i=1}^{n} p(E_i)$$

where $E_1 \cup E_2 \cup \ldots \cup E_n$ is the event which occurs if any of $E_1, \ldots, E_n$ occur, and which does not occur otherwise. Now in the standard treatment (for example Kolmogorov's axioms), proposition (2) (*finite additivity*) can be strengthened to give (*countable additivity*) which may be stated as

(2′) If $E_1, \ldots, E_n, \ldots$ are incompatible events,

$$p(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} p(E_i)$$

The stronger proposition (2′) does not follow from de Finetti's axioms, and de Finetti in fact regards the introduction of countable additivity as an illegitimate extension. This constitutes a most significant departure from the usual approach, and we shall examine it in detail in the next section.

Before turning to this matter, however, we should mention that Shimony in his [1955] introduced a stronger concept of coherence; now commonly called 'strict coherence'. *A*'s betting quotients are said to be coherent if it is impossible for *B* (his opponent) to choose the stakes *S* such that he wins whatever happens. A series of betting quotients would thus be coherent in a situation in which *A* might lose and could at best break even. Such a situation would be highly unsatisfactory for *A*, and therefore Shimony suggests that *A* should choose his betting quotients so that there is always a chance of his winning. If *A* obeys this condition his bets are said to be *strictly* coherent. Shimony's stronger condition seems at first sight highly plausible, but it has a rather curious consequence. If we adopt strict coherence, then (p. 141): 'Zero probability signifies impossibility'. This fact renders the condition unacceptable to de Finetti.

4 *Finite versus Countable Additivity.*

We shall begin by describing some of the consequences which follow from the replacement of countable by finite additivity, and will then go on to examine the reasons de Finetti gives for making the change. In the usual approach the distribution function $F(x)$ $(= p(X<x))$ tells us everything about the corresponding random variable X, so that we may replace the study of random variables by the study of their distribution functions. This ceases to be the case when we restrict

ourselves to finite additivity. A simple example will illustrate what is involved. Suppose the distribution function has a jump of magnitude p at x_0. In the standard theory this would mean that $prob(X = x_0) = p$. In the finite additivity case not only can we not infer the probability of $X = x_0$, but x_0 may not even be in the set of possible outcomes. To prove this let us begin by remarking that in the finite additivity case we can have uniform distributions over the integers by putting $prob(n) = p_n = 0$ for each n. This is not legitimate with countable additivity because we could deduce that $prob(Z) = \sum_{n=-\infty}^{\infty} p_n = 0$, where Z is the set of integers; and this would contradict the axiom that the sample space must have probability 1. In the finite additivity case we can no longer deduce from $p_n = 0$ for all n that $prob(Z) = 0$, and thus the assignment $p_n = 0$ for each n becomes legitimate.

Let us suppose (see p. 283) that the possible outcomes are $x_0 - 1$, $x_0 - \frac{1}{2}, \ldots$, $x_0 - 1/n, \ldots$, and we assign a uniform distribution to this sample space. The corresponding distribution function will have a jump of magnitude 1 at $X = x_0$, but x_0 is not even a possible outcome.

A second consequence of the restriction to finite additivity concerns the definition of 'strong convergence'. In the standard treatment, we say that a sequence of random variables $X_1, \ldots, X_n, \ldots$ converges to a random variable X in the strong sense iff

$$prob(\lim_{n=\infty} X_n = X) = 1$$

This definition no longer becomes appropriate if we have only finite additivity, and we must say instead that strong convergence occurs iff given $\epsilon > 0$ and $\theta > 0$, there is an N such that for K arbitrarily great and $n \geqslant N$.

$$prob(\bigcap_{k=0}^{K}(|X_{n+k} - X| < \epsilon)) > 1 - \theta$$

Already in his [1937] de Finetti used this definition in formulating and proving a version of the strong law of large numbers.

Perhaps an analogy will help to clarify the situation. The difficulties here are similar to those which give rise to the concept of *separable* stochastic processes (cf. Doob [1953], Ch. II, section 2, pp. 50–71). This latter problem crops up, when, assuming countable additivity, one tries to study simultaneously an uncountable number of random variables, *e.g.* X_t for $-\infty < t < \infty$. As Doob points out, one cannot then from

$$prob(X_t = 0) = 1 \quad -\infty < t < \infty$$

conclude that

$$prob(X_t = 0, -\infty < t < \infty) = 1$$

but only that

$prob(X_t = 0, t \in T) = 1$ where T is any finite or countable subset of the real line. Analogously de Finetti investigates difficulties which arise, when, assuming finite additivity, one studies countable collections of random variables, or of possible outcomes.

Such then, in brief outline, are some of the consequences of restricting oneself to finite additivity. But why should one adopt this restriction? De Finetti puts forward three principal arguments for doing so. First of all he points out that no one has given any reason for adopting countable additivity beyond saying that it is 'mathematically convenient' to do so. As he puts it (p. 141): 'No one has given a substantial justification of countable additivity (beyond saying that it is a "natural extension" of finite additivity: . . .)'. He quotes (p. 271), as typical of this attitude, J. F. Kingman and S. J. Taylor ([1966], p. 101): '. . . *there would not be much of analysis left* if one could not carry out at least sequential limiting operations' (de Finetti's italics). Against this, de Finetti's point is that the theory of probability is not a mere mathematical toy, but has a significant content. One should not therefore arbitrarily introduce axioms which give greater mathematical elegance unless these axioms can be justified in terms of the meaning given to probability. Now from the subjective point of view there is a justification for finite but not for countable additivity. De Finetti strengthens his case here by pointing out that the same holds in a philosophical account of probability which is quite opposed to his—von Mises's frequency theory (see p. 147). From von Mises's limiting frequency definition, we can derive finite but not countable additivity. Von Mises deals with this problem in his last work on the subject (von Mises [1963]) by arbitrarily postulating countable additivity. However a restriction to finite additivity would have accorded much better with his general philosophical position. I myself agree with de Finetti that we need a more substantial justification of countable additivity than that provided by 'mathematical convenience'. It might however be possible to provide such a justification in a philosophical theory different from the subjective or frequency views (for example in the propensity theory). I will not however pursue this question further.

As we have remarked already, a restriction to finite additivity enables us to introduce uniform distributions over the integers. De Finetti regards it as a counter-intuitive feature of the axiom of countable additivity that it prevents us from having such uniform distributions. After all, for any finite n however large, we can introduce a uniform distribution over the positive integers $1, 2, \ldots, n$ by setting $prob(i) = p_i = 1/n, i = 1, \ldots, n$. However if we postulate countable additivity, over the infinite collection of positive integers $1, 2, \ldots, n, \ldots$ we can only have what de Finetti terms (p. 145) 'extremely unbalanced distributions'. He explains his meaning here more fully on page 736 when he says that, with countable additivity, we have: 'to select from them (the countable class in question, *e.g.* the positive integers) a finite number to which to attribute a total probability of at least 99 per cent (leaving for all the others a total probability of less than 1 per cent; and I could have said 99.999 per cent and 0.001 per cent and worse)'. Another approach to this problem is provided by Rényi's concept of 'conditional probability space' (cf. his [1955]), for, in such spaces, we retain countable additivity and yet can, in a certain sense, define uniform distributions over the integers. De Finetti (p. 737) dismisses this as 'an ad-hoc device'.

De Finetti's third argument against countable additivity involves the claim (see pp. 271–4) that it inevitably takes us further than we may at times want to go, while preventing us from going as far as we might at other times wish. This can be best explained by means of an example. Suppose our sample space is the real interval [0,1] and we have assigned our subjective probabilities (betting quotients) to the subintervals. If we postulate countable additivity, we have thereby automatically assigned a probability to every Borel set. Yet we may not wish to bet on any of the more complicated such sets. Conversely countable additivity excludes us from assigning a probability to certain sets. It follows from an extension by Banach, Kuratowski and Ulam of a theorem of Vitali-Lebesgue, that we cannot, if we postulate countable additivity, extend the probability measure to all the subsets of the interval. Yet we *may* want to bet on any arbitrary set—no matter how complicated. De Finetti feels that no axiom should prevent us from doing so. This leads at once to finite additivity, for if we restrict ourselves to finite additivity, we can extend the probability measure to all the subsets of the interval.[1]

Such then are de Finetti's arguments against countable additivity. Whether or not they are convincing is a matter which each reader must evidently decide for himself, but all should agree that de Finetti has done a service of the greatest value in challenging a dogma in the foundations of probability, which is usually both unjustified and unquestioned. Moreover his investigations of the effects of replacing countable by finite additivity are obviously of very great mathematical interest—whatever axiom is finally adopted.

5 *Independence and the Laws of Large Numbers.*

A large part of the traditional mathematical theory of probability is concerned with the assumption of 'independence' and with the derivation of 'laws of large numbers' for sequences of independent random variables. It is therefore very interesting to ask how these notions are interpreted within the subjective theory.

De Finetti defines the stochastic independence of two events by the usual product formula (A and B are stochastically independent iff $P(A \cap B) = P(A)P(B)$). However in this formula the probabilities have a subjective meaning and are relative to a particular person. The same therefore holds for the notion of independence itself. As de Finetti himself puts it (p. 179):

> 'Even if, for brevity, one says at times that two events . . . *are* stochastically independent, it must be remembered that it is simply understood . . . "according to the opinion of the person who has selected the evaluation P".'

In his discussions of independence, de Finetti is constantly stressing that the assignment of stochastic independence (*i.e.* the assignment of subjective probabilities which satisfy the corresponding 'product' formulas) is in general, and perhaps always, unreasonable. Indeed he takes the laws of large numbers as demonstrating this very fact. As he puts it (p. 392):

> 'The principal practical application of the law of large numbers consists in showing how little it is realistic and practically reasonable to make the strict admission of equiprobability and stochastic independence . . .'

To explain the argument here, let us take the simplest case. Suppose we have a sequence of random variables $X_1, \ldots, X_n, \ldots$ each of which can take the values

---

[1] Dr. M. Machover kindly pointed out to me that this last argument of de Finetti's needs strengthening in view of a recent result of Solovay's. In his [1970] Solovay has shown that we can extend a measure on say the sub-intervals of [0,1] to all the subsets of [0,1] while retaining countable additivity provided that we weaken the axiom of choice. Moreover the weakened version of the axiom of choice used by Solovay is apparently sufficient for many of the results of classical mathematics normally proved using the ordinary axiom of choice. De Finetti has now to show, in order to render his argument satisfactory, that it is better to weaken countable to finite additivity rather than to weaken the axiom of choice used.

0 or 1. Suppose further that we postulate that they are independent and $prob(X_i = 1) = p$ for $i = 1, \ldots, n, \ldots$. The law of large numbers then tells us (speaking loosely) that for large n the frequency ratio of successes has a high probability of being near p. In other words, if we make such a probability assignment, we are committed to betting at very high odds of getting a frequency ratio of successes near p for large n. However we may well not want to commit ourselves to such a risky bet, and correspondingly should avoid the 'independence and equiprobability' assignment.

The need for avoiding assignments of independence within the subjective theory is made even clearer when de Finetti comes to discuss inductive reasoning in Chapter XI. It is part of his theory that an initial, or 'a priori', assignment of probabilities can never be abandoned as being 'falsified by experience'. It can only be modified by experience to give 'a posteriori' probabilities. The consequence of this for independence assignments are stated thus (pp. 576–7):

> 'And if I select initially an opinion based on independence, that is to say that I exclude once and for all the possibility of judging significant and adapted to modify my opinion any one of those results which could be denounced as "curious".
>
> ... To have the right to be influenced by experience it will therefore be necessary to have selected an initial opinion different from that of independence.'

This makes the situation admirably clear. Suppose we have a sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$. To judge them independent is equivalent, on de Finetti's account, to making the decision that, whatever values we discover $X_1, X_2, \ldots, X_{n-1}$ to have, we will not allow this to alter our opinion about the probability distribution of X_n. None but the most extreme dogmatist could, in my opinion, knowingly make such a decision. Thus independence is, practically speaking, excluded from the subjective theory. We must replace the concept of independence by another concept, and this is, of course, the concept of 'exchangeability' introduced by de Finetti himself in his early papers. We will discuss this concept in the next section, but we will first pause to describe by way of contrast, the role of independence within an objective account of probability.

To the objectivist[1] the claim that a sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ are independent is a conjecture about reality. We are not prevented from abandoning such an opinion as 'falsified by experience'. Quite on the contrary, we actually try to falsify such a claim by subjecting it to many and severe statistical tests. If such tests bring to light sufficiently 'curious results', we regard our original (initial) conjecture as having been disproved by experience, and try to replace it by a new hypothesis which may be quite different from our original one and not a mere modification. The laws of large numbers and other theorems of the mathematical calculus show, as de Finetti quite rightly stresses, that an assumption of independence entails certain consequences which we might be disinclined to bet on 'a priori'. For the subjectivist this means that we should be disinclined to postulate independence. For the (Popperian) objectivist, however, it means that we should be all the more inclined to conjecture independence, for such a conjecture is thereby shown to be significant (contentful). Of course, like any significant hypothesis, a hypothesis of independence may well be falsified by the tests we design and carry out, but, if it passes these tests, then we have succeeded in corroborating an interesting conjecture about reality.

---

[1] Or at least to the Popperian objectivist, cf. Popper [1957] where the objectivist points are forcibly made.

6 *Exchangeability: Objective Probabilities in a Subjective Theory.*

De Finetti deals with the notion of exchangeability mainly in Chapter XI of his book. However the treatment he provides is less full than that given in his original [1937], and I shall therefore draw my quotations in this section mainly from that paper. It does not seem to me that de Finetti has altered his viewpoint significantly on this question. So I do not believe this procedure is a distortion.

The question of exchangeability is closely connected with what I regard as the fundamental problem facing any subjective theory of probability. Most objectivists would not want to claim that there are objective probabilities in such situations as betting on horses, or on the results of elections, *etc.* Indeed von Mises specifically denied that his frequency theory of probability applied in such cases. The objectivists' claim is rather that in certain other areas—physics, life insurance, games of chance with coins, cards and dice, *etc.*—there are objective probabilities whose values can, in favourable circumstances be evaluated accurately. Such objective probabilities are analogous to other scientific quantities —forces, masses, electrical charges etc.—whose values can likewise be determined, in appropriate circumstances, to a certain degree of precision. It is a good piece of prima facie evidence in favour of the objective view that people do say things like 'The probability of getting heads with this bent coin is p' or 'The probability of this atom decaying radioactively in the next hour is q' where p,q are more or less accurately specified constants whose values, within the appropriate limits, are generally agreed. How is the existence of such probabilities to be explained on the personalistic 'degree of belief' point of view, according to which each probability judgment expresses the opinion of a particular individual? How, in effect, can we have apparently objective probabilities within a subjective theory?

De Finetti is of course well-aware of this problem and puts it like this ([1937], p. 152):

> 'It would not be difficult to admit that the subjectivistic explanation is the only one applicable in the case of practical predictions (sporting results, meteorological facts, political events, *etc.*) which are not ordinarily placed in the framework of the theory of probability, even in its broadest interpretation. On the other hand it will be more difficult to agree that this same explanation actually supplies rationale for the more scientific and profound value that is attributed to the notion of probability in certain classical domains.'

I now propose to analyse and criticise de Finetti's solution to this problem—a solution, which, as we shall see, depends critically on the notion of exchangeability. It will be easier to discuss the matter in terms of a particular example. We will first describe this from the objective point of view and then examine how de Finetti proposes to treat it. The example is this. We have two men M and N. N has a coin which looks fair but for which in fact $prob(heads) = \frac{2}{3}$, and not $\frac{1}{2}$. N tosses the coin a large number of times, and M initially supposes that the coin is fair, *i.e.* $prob(heads) = \frac{1}{2}$. However after a while he observes that there has been an excessive preponderance of heads. He therefore regards his initial hypothesis *prob*(*heads*) $= \frac{1}{2}$ as falsified, and adopts the hypothesis *prob*(*heads*) $= \frac{2}{3}$ instead. This agrees with his subsequent observations and he finally concludes that *prob*(*heads*) really is $\frac{2}{3}$ or at least that this hypothesis is well-corroborated. So much for the objective description of this example. We must now examine how de Finetti would deal with it.

On his approach M has to consider at what odds he would bet if he was forced to do so by N. He decides that his initial betting quotient on heads should be $\frac{1}{2}$. He also decides *a priori* what conditional bets he would make on future tosses. A conditional bet on B given A at rate p is defined thus. If A occurs, we bet on B at rate p. If A does not occur, we call off the bet. In the particular case we are considering, M has to propose conditional betting quotients on the result E_{n+1} say of the $n+1$th toss given all possible results of the first n tosses. These various betting quotients must satisfy the further condition of exchangeability. In our present case this means that the *a priori* betting quotient for getting a particular sequence of r results (Π_r say) on tosses $i_1, i_2, \ldots, i_r$ depends perhaps on r but not on $i_1, \ldots, i_r$ for any $r \geqslant 1$. Intuitively this amounts to saying that we regard any particular group of r tosses as 'exchangeable' with any other—an apparently reasonable requirement.

The various betting quotients introduced must of course be coherent. Hence they must satisfy the axioms of probability and in particular Bayes's theorem. Therefore using fairly standard methods of reasoning de Finetti shows that whatever (within broad limits of course) *a priori* probability $p(E_{n+1})$ we introduce for getting say heads on the $n+1$th toss, the conditional probability $p(E_{n+1} \mid A)$ will tend to r/n where A states that there were r heads on the first n tosses. De Finetti assumes that, if we observe A, we will bet on E_{n+1} at the betting quotient $p(E_{n+1} \mid A)$. Thus whatever our 'a priori' betting quotients, our 'a posteriori' betting quotients will tend more and more to the observed frequencies. The objectivist will incorrectly suppose that we are discovering the 'true' probability. But, according to de Finetti, all that is happening is this. Two men starting with different personal degrees of belief are gradually forced by the constraints of coherence and exchangeability to adopt degrees of belief which become increasingly similar and approximate more and more closely to the observed frequency.

It is worth stressing two assumptions on which the above argument depends. First of all (as we mentioned in the previous section) a certain 'a priori' distribution must be adopted before the first toss, and this distribution is never given up but rather replaced by an 'a posteriori' distribution. That is to say that we can never state that our *a priori* probabilities were wrong and we should instead have adopted such and such different *a priori* probabilities. Secondly it is assumed that when we bet on E_{n+1} having observed result A on the first n tosses, we use our original conditional betting quotient $p(E_{n+1} \mid A)$. De Finetti himself lays emphasis on these two points in the following passage ([1937], p. 146):

> 'Whatever be the influence of observation on predictions of the future, it never implies and never signifies that we *correct* the primitive evaluation of the probability $p(E_{n+1})$ after it has been *disproved* by experience and substitute for it another $p^*(E_{n+1})$ which *conforms* to that experience and is there-

> fore probably *closer to the real probability*; on the contrary, it manifests itself solely in the sense that when experience teaches us the result A on the first n trials, our judgement will be expressed by the probability $p(E_{n+1} \mid A)$, i.e. that which our original opinion would already attribute to the event E_{n+1} considered as conditioned on the outcome A. Nothing of this initial opinion is repudiated or corrected; it is not the function p which has been modified (replaced by another p^*), but rather the argument E_{n+1} which has been replaced by $E_{n+1} \mid A$, and this is just to remain faithful to our original opinion (as manifested in the choice of the function p) and coherent in our judgement that our predictions vary when a change takes place in the known circumstances.'

The criticism we shall now give of de Finetti's argument was first clearly formulated by Hacking in his [1967], Sections 2 and 3, though there are hints of it (as Hacking points out) in Suppes [1966], Section 4. Hacking attacks the second of the assumptions which we listed above. He observes that there is no need, so far as coherence and exchangeability are concerned, to adopt the conditional betting quotient $p(E_{n+1} \mid A)$ as one's betting rate on E_{n+1} once A has actually occurred. All the initial betting quotients $p(E_1)$, $p(E_{n+1})$, $p(E_{n+1} \mid A)$, *etc.* refer to actual or hypothetical bets made before the result of the first toss is observed. Coherence and exchangeability require that they should obey certain conditions. But suppose one or more results are known. We are now in a different betting situation and the previous constraints no longer apply. In particular if the result A of the first n tosses is known, we could coherently adopt any betting quotient (p say) on E_{n+1} provided we adopt $1-p$ on E_{n+1}. We do not have to put $p = p(E_{n+1} \mid A)$.

Once this is pointed out, we can easily find situations in which someone might quite reasonably not want to set $p = p(E_{n+1} \mid A)$. Suppose again that N is forcing M to bet. N shows M, from a distance a coin which looks a little bent. He tells him that the coin is biased, but refuses to say in which direction it is biased. M has to make the usual conditional and absolute bets. As the game proceeds he is told the results of the tosses but is not given any more clues as to the direction of the bias.

Before the first toss M reasons thus: 'The coin is definitely biased, but I do not know how much or in what direction. Thus for me any of the results 0, 1, . . ., n heads on the first n tosses are equally probable.' This reasoning leads him to adopt, *a priori*, the Laplacean distribution, and to choose all the necessary betting quotients accordingly. In particular, as de Finetti in fact points out ([1937], pp. 143–5), his argument gives $p_r^{(n)} = r+1/n+2$, where $p_r^{(n)}$ is the probability of heads on the $n+1$ th toss given r heads on the first n tosses. So in particular $p_1^{(1)} = \frac{2}{3}$, *i.e.* if one head is observed M should change his betting quotient on heads from $\frac{1}{2}$ to $\frac{2}{3}$.

Suppose next that on the first toss the result is indeed heads. Suppose further that, while the first toss is taking place, M has continued to think about the problem. He now reasons thus: 'The coin did not after all look *very* bent. It is much more likely to have a small rather than a large bias. Thus I was wrong to consider each of the results r heads on n tosses as equally probable. I should have given a greater probability to results for which $r/n \doteqdot \frac{1}{2}$.' As a result of these new thoughts, M chooses for $p_1^{(1)}$ after the first toss, not $\frac{2}{3}$, but a value between $\frac{2}{3}$ and

$\frac{1}{2}$ and much nearer $\frac{1}{3}$. His behaviour, as I have described it, seems perfectly rational throughout. Yet he does not change his betting quotients in accordance with Bayesian conditionalisation.

This example can of course be generalised. We can imagine someone who throughout a whole series of tosses, continues to rethink the problem and improve his original analysis of the situation. As a consequence, he might rarely or never change his betting quotients according to Bayesian conditionalisation. Yet I would regard such a person as highly rational—more rational indeed than someone who remained faithful to his original conditional betting quotients.

Another possibility is this. We may *a priori* consider the event A to be very unlikely to occur, and adopt the conditional betting quotient $p(E_{n+1} \mid A)$ *not* because we considered this a reasonable betting quotient given that A has actually occurred but merely in order to be coherent with betting situations we thought more likely to occur. If A actually did occur our whole outlook might alter, and we would want to adopt a betting quotient on E_{n+1} different from $p(E_{n+1} \mid A)$.

It is interesting to examine Ramsey's attitude to the present problem, for in fact he contradicts himself on this point. In his [1926], page 180, he writes: 'The degree of belief in p given q is measured thus.' The usual definition of a conditional betting quotient is given and he continues: 'This is not the same as the degree to which he would believe p, if he believed q for certain; for knowledge of q might for psychological reasons profoundly alter his whole system of beliefs.' This seems to be, in effect, a statement of Hacking's argument. Yet in the same article ([1926], p. 192) we find:

> 'We have therefore to explain how exactly the observation should modify my degrees of belief; obviously if p is the fact observed, my degree of belief in q after the observation should be equal to my degree of belief in q given p before, ... When my degrees of belief change in this way we can say that they have been changed consistently by my observation.'

However to return to de Finetti, we may supplement Hacking's fundamental criticism by some further observations. First of all de Finetti envisages a complicated system of absolute and conditional bets which are made, or at least considered, *before* the first toss. Now this is an unrealistic assumption when applied to actual betting situations. Consider for example gamblers at Monte Carlo. They bet on a spin of the wheel. The result is observed, and they then go on to make a different set of bets on the next spin *etc.* No conditional bets are ever made or even considered. Yet all accept the 'objective' odds as being the correct ones. (I neglect the fact that the proprietors in fact offer unfair odds in order to increase their profits.) It thus seems that objective probabilities can be established quite apart from a consideration of conditional bets and Bayesian conditionalisation. Moreover in such betting situations coherence only gives us that the sum of our probabilities on the exhaustive and exclusive set of outcomes must be 1. It does not further determine the values that we should adopt.

This argument is even more striking in the case of probabilities in physics. Any standard textbook of atomic physics which discusses radioactive elements will give values for their atomic weights, atomic numbers, *etc.*, and *also* for the probability of their disintegrating in a given time period. These probabilities appear to be objective physical constants like atomic weights, *etc.*, and, like these last quantities, their values are determined by a combination of theory and experiment. In determining the values of such probabilities no bets at all are made or even considered. Yet all competent physicists interested in the matter agree on certain standard values—just as they agree on the values of atomic weights, *etc.*

We see then that the kinds of bets envisaged by the subjective theory are unrealistic. We can further question whether such bets would always accurately represent an individual's degree of belief. To see this suppose as usual that M is betting with N.[1] N asks M to give an *a priori* betting quotient for the 1,001st toss of a certain coin, but not to specify further betting quotients. M examines the coin and since it looks fair gives the betting quotient $\frac{1}{2}$. N specifies a positive stake S for the result heads on the 1,001st toss, *i.e.* if heads occurs on that toss M gains $\frac{1}{2}S$; if tails occurs, N gains $\frac{1}{2}S$. The coin is now tossed 1,000 times and a random sequence with frequency ratio of heads of about $\frac{1}{4}$ is observed. Naturally if *now* asked to give a betting quotient which represented his degree of belief in heads on the 1,001st toss, M would choose $\frac{1}{4}$. But now N can win money off M by setting as stake on the second bet $-S$. For in the event of heads occurring M's total gains on the two bets are $\frac{1}{2}S - \frac{3}{4}S = -\frac{1}{4}S$, and if tails occurs $-\frac{1}{2}S + \frac{1}{4}S = -\frac{1}{4}S$. So he loses in both cases. Indeed to avoid such a Dutch book, M has to choose $p(\textit{heads}) \geqslant \frac{1}{2}$.

What has gone wrong here? It is clear that, in the example cited, M has to choose as betting quotient *not* something that represents his true degree of belief, but something that will prevent N winning off him, granted that certain bets have previously been made. So for M's betting quotient to represent his belief, we have to neglect the effect of bets made previously. But if we neglect previous bets, then we need not change our betting rates according to Bayesian conditionalisation.

My last criticism of de Finetti's argument again concerns the assumption it involves that we do not alter our *a priori* assessment of the situation in any radical fashion but only modify it by Bayesian conditioning. In fact however there are cases where we may want to change our initial formulation of the problem completely, and de Finetti himself even mentions such a case. To explain this we shall have to give an account of his very ingenious result concerned with exchangeable events which now goes by the name of de Finetti's theorem.

Let $X_1, \ldots, X_n, \ldots$ be exchangeable random variables and let $Y_n = 1/n\sum_{k=1}^{n} X_k$, then de Finetti first shows ([1937], p. 126) that the distribution $\Phi_n(\xi) = p(Y_n \leqslant \xi)$ tends to a limit $\Phi(\xi)$ as $n \to \infty$. He goes on to say ([1937], p. 128–9):

'Indeed, let $p_\xi(E)$ be the probability attributed to the generic event E when the events $E_1, E_2, \ldots, E_n, \ldots$ are considered independent and equally probable with probability ξ; the probability $p(E)$ of the same generic event, the E_i being exchangeable events with the limiting distribution $\Phi(\xi)$, is

$$p(E) = \int_0^1 p_\xi(E)\, d\,\Phi(\xi).$$

---

[1] The following example was suggested to me by discussions with Mr J. Dorling. See, in this context, his [1972], p. 148, footnote 4.

Thus the system of exchangeable events with limit distribution $\Phi(\xi)$ corresponds to a certain situation which might be envisaged by an objectivist; namely one in which he supposes the events $E_1, \ldots, E_n, \ldots$ to be independent with true but unknown probability $p_\xi(E)$ where ξ is a random variable lying between 0 and 1 with distribution function $\Phi(\xi)$. Similarly de Finetti thinks ([1937], p. 146, footnote 4):

> 'One could in the first place consider the case of classes of events which can be grouped into Markov "chains" of order 1,2, . . ., *m*, . . ., in the same way in which classes of exchangeable events can be related to classes of equiprobable and independent events.'

Let us call such classes of events, Markov-exchangeable events. These according to de Finetti would constitute a complication and extension of his theory, but do not cause any fundamental problem. Indeed he says ([1937], p. 145):

> 'One cannot exclude completely *a priori* the influence of the order of events, ... There would then be a number of degrees of freedom and much more complication, but nothing would be changed in the setting up and the conception of the problem, . . ., before we restricted our demonstration to the case of exchangeable events, . . .'

But now suppose we have the situation which would be described objectively as follows. We have a series of events produced by some physical process. We first think that they are independent and that the various possible outcomes have constant probabilities. Then, after examining some results we conclude that the independence assumption is inadequate and decide instead that we are dealing with a Markov chain. This Markov chain hypothesis in fact agrees with the further results obtained. Such a situation could not, I claim, be dealt with in de Finetti's system. Corresponding to the initial independence assumption, he would make an exchangeability assumption. But this a priori assumption once made can never, as he himself says, be revoked. If we make a Markov-exchangeability assumption a priori we are all right, but once we have made the exchangeability assumption we can never change over to Markov-exchangeability in the light of the evidence. This seems plainly unsatisfactory. On the other hand if we allow a change to Markov-exchangeability in mid-channel (so to speak), then it becomes false that all changes of probabilities are by Bayesian conditionalisation and the whole basis of de Finetti's argument is overthrown.

I conclude that de Finetti's argument, for all its ingenuity, does not succeed in explaining, from his point of view, the existence of apparently objective probabilities, *i.e.* the existence of probabilities whose values are in fact generally agreed.

7 *Inductive Reasoning.*

One of de Finetti's boldest claims is that his theory solves the traditional philosophical problem of induction, and gives a general account of inductive reasoning. He sets out his arguments on this subject at the beginning of Chapter XI. The problem of induction can be put like this. Suppose we have a universal scientific hypothesis *h*, *e.g.* all ravens are black, and some evidence *e* for *h*, *e.g.* reports of the observation of *n* ravens all of which turned out to be black. Now *h* does not follow logically from *e*. What reason have we therefore for believing *h* on the basis of *e*? In particular why should we hold that the next raven we meet is more likely to be black than white or red? De Finetti's answer is that although *h* does not follow logically from *e* it is rendered probable by *e*. More precisely the probability $p(h,e)$ of *h* given *e* is higher than the prior probability $p(h)$ of *h*, both these probabilities being interpreted in the usual subjective sense. Inductive reasoning consists in replacing $p(h)$ by $p(h,e)$ where the latter is calculated using Bayes's theorem.

De Finetti intends his account of inductive reasoning to apply not only to the scientific case just mentioned, but to all other cases as well. He lists these on page 559 where he says:

> 'The field of induction is extended in every direction and to all levels: from the value of the arguments for and against the trustworthiness of different scientific theories or for and against the guilt of this or that suspect in the case of a crime, to the methods for establishing on the basis of observation the conditions for certain types of insurance and to those for obtaining adequately precise values of quantities by means of measurements which are inevitably imprecise.'

He stresses that his account is the same in all these cases (p. 569):

> 'A fact which needs stressing is that the explanation offered is unique, valid without distinction for all inductive reasoning.'

The explanation in question is that, in inductive reasoning, we replace prior (subjective) probabilities by posterior (subjective) probabilities in accordance with Bayes's theorem. As de Finetti himself puts it (p. 567):

> '. . . all that can be said regarding "inductive reasoning" . . ., essentially, reduces . . . to Bayes's theorem.'

Of course, he does not deny that there may be cases where we do not and would not wish to evaluate the probabilities concerned precisely. However these cases are still qualitative applications of the same fundamental reasoning. Thus in discussing the complicated processes by which new scientific concepts and theories are developed, he says (p. 559):

> 'Essentially we still have an analysis of the knowledge situation made on the basis of Bayes's theorem, except that, in such broad and vague situations, it would be impossible to think of an application of it which is not merely qualitative.'

There seems to me to be an obvious objection to this account of de Finetti's. Suppose we are asked to bet on a (universal) scientific hypothesis, *e.g.* all ravens are black. The only sensible betting quotient to choose is 0, because we might lose our bet (if a non-black raven is discovered) but we can never win it; for it can never be established that all ravens are black, but only that all so far observed are black. Thus the prior subjective probability $p(h)$ of *h* will be 0, and so, by Bayes's Theorem the posterior subjective probability $p(h,e)$ will also be 0. Yet this contradicts our intuition that our confidence in *h* should increase with increasing evidence in its favour.

One way out of this difficulty consists in considering 'the probability of a universal hypothesis' as a convenient shorthand for 'the probability of certain associated facts'. Thus instead of considering the probability of all ravens being black, we might consider the probability that the next hundred observed ravens will be black. De Finetti advocates a view of this sort in his [1971]. In discussing a case where we have a set of possible laws each specified by a value of a parameter θ (a typical statistical example), he says ([1971], p. 90): 'Such a point of view takes as "principal objective" not the distribution of the parameter (that is: opinions regarding the "laws") but rather the "predictive" probability distribution of the related "facts".' This approach naturally leads to an 'instrumentalist' treatment of scientific laws. They are considered not as statements which might be true or false, probable or improbable, known or unknown, but as heuristic rules of inference leading from facts observed in the past to others which will perhaps be observed in the future. As de Finetti puts it ([1971], p. 89):

> 'The formulation of a theory, or of a law is a link in the chain . . . of the mental process by which we pass from the observation of passed facts to the prevision of future facts.'

There are a number of difficulties involved in this account. To begin with it needs elaboration. If we are to reduce the probability of theories to the probability of certain associated facts, some indication should be given of how we choose the associated facts in various cases. Secondly (and more seriously) there is the point that facts presuppose theories or, as it is sometimes put, that all facts are theory laden.[1] For example a typical 'fact' of modern science might be that such and such a substance, when heated, emitted light of such and such a wavelength. This fact presupposes certain theories about the nature of light. In the absence of such theories it would be meaningless to speak of the 'wavelength' of light. But if facts presuppose theories, how can we reduce theories to facts?

A possible answer would be this. Instead of explaining the concept of 'wavelength of light' by means of theories, we give an operationalist definition of it in terms of observables. Further these observables are described in a 'theory free' language. By using such definitions, we can transform theory laden into theory free facts, and then reduce theories to such facts. There has been a long history of attempts to carry out such a programme, but none of them has to date met with success. Yet it seems to me that de Finetti needs to bring such a programme to successful completion before he can justifiably claim to have solved the problem of induction.[2]

8 *Mathematical Aspects.*

Turning now to the mathematical parts of de Finetti's book, we must remark straight away that he has a most unusual mathematical style. In a standard textbook one normally finds proofs and theorems presented as if God-given and with 'remorseless logic'. There is generally little or no attempt to explain to the reader the underlying idea of the proof or to give an indication of how it came to be discovered. De Finetti criticises, quite rightly in my opinion, this over-formalistic approach to mathematics and tries instead in his own treatment to make things

[1] Duhem seems to have been the first to make this very important point ([1906], Part II, ch. 4), but it has been much discussed by recent philosophers of science.

[2] See however Dorling [1972] for arguments in favour of a personalist approach to inductive inference. See also Lakatos [1968] for a fuller discussion of the subjective and related approaches to the problem of induction.

intuitively clear.[1] De Finetti also inveighs against the tendency to use the latest sophisticated methods even for simple problems and tries instead to deal with simple problems by correspondingly simple methods, reserving more advanced methods for occasions where they are really needed.

The principal mathematical discussions of the book are found in Chapter 8 which treats of processes with independent increments. The choice of this particular subject need not surprise us, since de Finetti himself initiated its study in his classic [1929]. In this area I feel that his more intuitive approach is highly successful. A good example is provided by his derivation of the canonical formula for infinitely divisible distributions. This formula is usually obtained (for example in Doob [1953], Ch. III, Section 4, pp. 132–5) by analytic methods which are short but give little insight into the real meaning of the result. De Finetti on the other hand in his admittedly much longer derivation (pp. 433–49) shows how we can reach the result by putting together various stochastic processes. This not only shows how the formula came to be discovered, but also explains the intuitive meaning of the various terms, and the need for introducing certain correcting factors.

The only criticism I would like to make of de Finetti here is that he is perhaps rather unduly modest about his own original contributions to the subject. The full story is told in Gnedenko and Kolmogorov [1949] and runs as follows. In his [1929], already mentioned, de Finetti began the study of processes with independent increments and introduced (implicitly) the concept of an 'infinitely divisible distribution'. This concept was formulated explicitly, though not given a name, in his [1930], where he also attempted to give a mathematical characterisation of infinitely divisible distributions. His results were improved by Kolmogorov in his [1932], where the canonical representation appears for the first time, but is demonstrated only in the case of finite variance. Then the formula normally quoted today was produced by P. Levy, working independently of de Finetti and Kolmogorov, in his [1934–5], and finally Khintchine in his [1937], and Feller in his [1939] gave shorter analytic proofs of the same result. De Finetti (p. 448) mentions the work of Khintchine and Feller, but omits to mention his own work of which theirs was a development.

Although de Finetti's intuitive approach seems to me highly successful in some instances, there are others in which (again for me personally) it is less satisfactory. This is so particularly in some of the earlier chapters where de Finetti replaces the usual analytic proofs by proofs involving n-dimensional geometry. For those whose geometrical intuition is anything comparable to de Finetti's, I am sure that these proofs will really give great insight into the problem; but perhaps many, like myself, will feel that the standard approach is easier here.

Apart from his main mathematical discussions, de Finetti introduces many interesting asides. It is standard to introduce (cf. Section 4) the indicator X_E of an event E as the random number which is 1 if E is the case and 0 otherwise. De Finetti goes further and suggests that we identify an event with its indicator. This means that logical operations on events become arithmetical ones (not $E = 1-E$, $E_1 \wedge E_2 = E_1E_2$, $E_1 \vee E_2 = E_1+E_2-E_1E_2$, . . .). There is a further interesting consequence. We can consider $(0 \leqslant x \leqslant 1)$ as the event (or the condition) that x lies between 0 and 1 inclusive, and hence as the corresponding random number. Thus instead of writing (see p. 23)

---

[1] For a similar criticism of the over-formalistic approach of mathematical text-books, cf. Lakatos [1963–4].

$$f(x) = \begin{cases} = 0 & \text{for } x < 0 \\ = g(x) & \text{for } 0 \leqslant x \leqslant 1 \\ = 0 & \text{for } x > 1 \end{cases}$$

we can simply write

$$f(x) = g(x)\ (0 \leqslant x \leqslant 1)$$

This shorthand proves useful on many occasions. It is moreover only one of several notational devices which de Finetti proposes.

There is another way in which de Finetti's treatment—even of quite elementary questions—proves more interesting than a standard one. There are many occasions on which a point is made by means of a standard example. On such occasions one usually finds in de Finetti some extra remark or generalisation of this example. A simple instance of this is found on page 176–7. De Finetti defines independence for n events in the usual way and makes the usual point (by means of the usual example) to the effect that pair-wise independence is not in general sufficient for independence. However he then, unusually, proceeds at once to generalise this example, and show that we can have independence up to a certain order m but not beyond.

I cannot conclude this review without making some mention of de Finetti's style in general (as opposed to mathematical style). How often do we find books on mathematics and philosophy written in a dull and prosaic style that weighs heavily on the reader. In contrast to this, de Finetti's prose sparkles with jokes, analogies and interesting asides, as well as abounding in allusions to and quotations from a wide variety of authors: Shakespeare, Dante, Caesar, ... With this feature in addition to the tremendous intellectual interest of its contents, the book was for me a very great pleasure to read, even though, as an objectivist, I cannot agree with de Finetti's conclusion that the subjective theory enables one to escape '*fuori del pelago alla riva*' (from the sea to the shore). I would rather see the subjective theory itself as the '*acqua perigliosa*' (dangerous water).

D. A. GILLIES
Chelsea College of Science and Technology

REFERENCES

Doob, J. L. [1953]: *Stochastic Processes.*

Dorling, J. [1972]: 'Bayesianism and the Rationality of Scientific Inference', *British Journal for the Philosophy of Science*, **23**, pp. 181–90.

Duhem, P. [1906]: *La Théorie Physique, son Objet et sa Structure.* (The reference is to the English translation of the second, 1914, edition: *The Aim and Structure of Physical Theory*, 1954).

Feller, W. [1939]: 'Neuer Beweis für die Kolmogoroff-Lévysche Charakterisierung der unbeschränkt teilbaren Verteilungsfunktionen', *Bulletin International Jugoslvenska Akademija, Classes des Sciences Mathématiques et Naturelles*, **32**, pp. 106–14.

de Finetti, B. [1929]: 'Sulle Funzioni a Incremento Aleatorio', *Rendiconti dell' Accademia Nazionale dei Lincei, Classe di Scienze Fisiche Matematiche e Naturali*, **10**, pp. 163–8, 325–9, 325–9, 548–53.

de Finetti, B. [1930]: 'Le Funzioni Caratteristiche di Legge Instantanea', *Rendiconti dell' Accademia Nazionale dei Lincei, Classe di Scienze Fisiche Matematiche e Naturali*, **12**, pp. 278–82.

de Finetti, B. [1937]: 'La Prévision: ses Lois Logiques, ses Sources Subjectives'. *Annales de l'Institut Henri Poincaré*, **1**, pp. 1–68. (Reprinted in English in H. E. Kyburg and H. Smokler (*eds.*): *Studies in Subjective Probability*, 1964, pp. 93–158.

de Finetti B. [1971]: 'Probabilità di una Teoria, e Probabilità dei Fatti', in *Studi di Probabilità, Statistica e Ricerca Opercitiva in onore di Giuseppe Pompilj.*

Gnedenko, B. V. and Kolmogorov, A. W. [1949]: *Limit Distributions for Sums of Independent Random Variables* (English edition, 1954).

Hacking, I. [1967]: 'Slightly More Realistic Personal Probability'. *Philosophy of Science*, **34**, pp. 311–25.

Keynes, J. M. [1920]: *A Treatise on Probability.*

Khintchine, A. Ya. [1937]: 'Deduction Nouvelle d'une Formule de M. Paul Levy', *Bulletin de l'Université d'État à Moscou*, Section **A.1**, pp. 1–5.

Kingman, J. F. and Taylor, S. J. [1966]: *Introduction to Measure and Probability.*

Kolmogorov, A. N. [1932]: 'Sulla Forma Generale di un Processo Stocastico Omogeneo (Un Problema di Bruno di Finetti)', *Atti dell' Accademia Nazionale dei Lincei. Rendiconti della Classe di Scienze Fisiche Matematiche e Naturali*, **15**, pp. 805–8, 866–9.

Lakatos, I. [1963–4]: 'Proofs and Refutations', *British Journal for the Philosophy of Science*, **14**, pp. 1–25, 120–39, 221–43, 296–432.

Lakatos, I. [1968]: 'Changes in the Problem of Inductive Logic', in I. Lakatos (*ed.*): *The Problem of Inductive Logic*, pp. 315–417.

Levy, P. [1934–5]: 'Sur les Integrales dont les Élements sont des Variables Aléatoires Indépendantes', *Annali della Reale Scuola Normale Superiore di Pisa*, **3**, pp. 337–66; **4**, pp. 217–18.

von Mises, R. [1963]: *Mathematical Theory of Probability and Statistics.*

Popper, K. R. [1957]: 'Probability Magic, or Knowledge out of Ignorance', *Dialectica*, **11**, pp. 354–73.

Ramsey, F P. [1926]: 'Truth and Probability', in F. P. Ramsey, *The Foundations of Mathematics*, 1931, pp. 156–98.

Rényi, A. [1955]: 'On a New Axiomatic Theory of Probability', *Acta Mathematica Academiae Scientiarum Naturalium Hungaricae*, **6**, pp. 285–335.

Shimony, A. [1955]: 'Coherence and the Axioms of Confirmation', *Journal of Symbolic Logic*, **20**, pp. 1–23.

Solovay, R. M. [1970]: 'A Model of Set Theory in which Every Set of Reals is Lebesgue Measurable', *Annals of Mathematics*, **92**, pp. 1–56.

Suppes, P. [1966]: 'Probabilistic Inference and the Concept of Total Evidence'. In J. Hintikka and P. Suppes (*eds.*): *Aspects of Inductive Logic*, pp. 49–65.

THE PLAIN MAN'S GUIDE TO PROBABILITY

Whatever effect the authoritative dismissal by Kendall and Stuart of what could be called the philosophy of probability might have had within their discipline—'In statistics it is a mark of immaturity to argue overmuch about the foundations of probability' (Kendall and Stuart [1951], vol. I, p. 180)—it has not had much outside. This book[1] contains a new, frequently obscure but always interesting contribution. It subscribes, moreover, to a philosophical credo which has tended to be inimical to technical display, holding on the contrary that a respectful attention to the vagaries of ordinary language affords the only key to the solution of philosophical problems. Here, however, there are all the signs of an informed and sincere concern with the mathematical basis and scientific applications of the theory.

---

[1] Lucas, J. R. [1970]: *The Concept of Probability*. Oxford: Oxford University Press. £2.10. Pp. xx+403.

Now among the stages in the development of any interpretative theory of probability are or should be the following:

(1) the axioms of probability should be shown to be satisfied

(2) certain informal desiderata postulated by the theory should be clearly achieved.

As to (1), each theory tends only to satisfy versions of these axioms, but this is usually no difficulty. As to (2) the desiderata vary enormously with the theory in question, and a useful way to classify theories, much better than the coarse and ambiguous partition into logical-frequency-subjective, is by listing the conditions they attempt to satisfy. Consider the following set

(i) the probability of an event A on repeatable and independent trials during which the probability remains constant is $\approx \frac{n(A)}{n}$ for large n:

(ii) that probability assignments be objective, *i.e.* their truth conditions do not involve reference to the mental state of any individual person or persons;

(iii) that they be subject to possible amendment in the light of experience or report—in short that they be testable;

(iv) that they furnish grounds for action, *e.g.* in betting. In other words $p(A)$ gives the correct odds (objective odds if (iii) is satisfied) on A occurring at the next trial, if A is such an event as in (i).

Among extant theories only the combination of a frequentist and propensity approach like Popper's (Popper [1959]) seems to satisfy all (i)–(iv). (This approach is unfortunately not discussed or even mentioned by Mr Lucas, and a great demerit of his book is its tendency to parochialism—Austin's and Toulmin's are the great names, it seems, and behind these the shadowy eminence of Wittgenstein is discernible—issuing in a rather cursory appraisal of non-Oxonian developments later than about 1935). So, ostensibly, does that expounded within this book. It should not be inferred that Mr Lucas has plagiarised Popper, for he postulates in effect a fifth desideratum which distinguishes him sharply from the entire frequentist tradition commencing with Venn. The condition is, briefly, that all everyday probabilistic judgments ('there is a probability greater than one half that Nixon will be re-elected') should be considered meaningful within the theory—should be, as it were, well-formed formulas. Now to find a theory of such apparent scope one would probably have to go back to Laplace and before him to Bernoulli (the subsequent discussion of Bernoulli's theorem will show this to be a rather apt parallel); indeed the appearance of a form of the principle of indifference in assigning probabilities *a priori*, and of Bayes's theorem in estimating probabilistic parameters *a posteriori*, give a character to this book that is almost more eighteenth than twentieth century.

The combination of these five desiderata is unorthodox; the presence of the fifth arises directly from what was Toulmin's, in the second chapter of his [1958], and is now the author's concern, namely to illuminate the notion of probability appealed to in 'ordinary' discourse (it seems to be axiomatic in this approach that there cannot be more than one). Toulmin's analysis ceased before any connection was made with the mathematical theory or even the axioms; indeed it is not at all obvious that such could exist, principally because there seems no good reason to suppose that the objects of the mathematical theory and whatever falls under the 'ordinary' notion of probability come from the same stable at all, or that they ought to be regarded as doing so. This book attempts to show that they do.

Primary, then, is the 'ordinary' concept of probability, and this governs the development of the author's theory. To discover why this restriction should be imposed at all one need only check back the references to Toulmin, and from Toulmin to Austin, whose influence throughout Toulmin's [1958] is apparent, even oppressively so, and to recall the celebrated remark of Austin that 'our common stock of words embodies all the distinctions men have found worth drawing, and the connections they have found worth making, in the lifetime of many generations; they are surely likely to be more numerous, more sound, since they have stood up to the long test of the survival of the fittest, and more subtle at least in all ordinary and reasonable practical matters . . .' and so on. The concept of probability is, by implication, not to be found in the work of von Mises, Kolmogorov or de Finetti, say.

Then what is the 'ordinary' concept of probability? Austin's example is followed by the author (' "*In vino veritas*", says Austin, "but in a sober symposium only *verum*" '[1]), in elucidating the noun by looking at the use of the adjective—here, 'probable'. The analysis proceeds much as did Toulmin's, in attending to the function of the word in 'ordinary' usage. Probability so elucidated turns out to be a 'modal' concept (attempts to assign it a *designatum*, by subjectivists, for example, are strongly reprobated) and indicates the extent to which assent to propositions should be granted. It is a guide of life, but a 'guarded guide', to quote the author. He is not averse to calling what he has gleaned from ordinary discourse a theory of rational belief: 'If we took the word "belief" in the sense in which it means more or less the same as the word "proposition", to say that probability was the degree of rational rightness of belief would be unexceptionable.'[2] The way is open to the mathematical theory by defining degrees of probability to be degrees of this quantity. Formally, 'the concept of probability' is a function from the formulas (though not all, as we shall see) of a formal language—in this case a system of many-sorted logic due to T. F. Smiley—into the closed unit interval. As we usually amend our estimates of probability, with increasing experience, and as this is a theory of absolute probabilities, conditional probabilities being defined subsequently, desideratum (iii) follows naturally enough. (i) is an adequacy requirement of practically any theory, and (ii) is ostensibly satisfied by this being a theory of rationality. (iv) arises because probability, 'ordinarily used', is a guide of life.

The first major task is to show that analogues of Kolmogorov's axioms follow within this theory. Actually, Mr Lucas speaks of justifying the axioms[3] as if they were of no wider significance than that which consists in being derivable from his theory; however, it is only by the introduction of highly questionable, even *ad hoc*, measures that they are in fact obtained. The first of these is his identification of the formal theory with, in effect, a many-valued logic. This enables him to identify extremal probabilities with truth and falsehood, and intermediate probabilities with intermediate truth values. In addition, he demands that this set of values be Boolean under conjunction, disjunction and negation. He has two arguments for the first stipulation. One is that 'the traditional calculus of probabilities [connects] probabilities with truth'[4]; this depends on what you call tradition. The other is elicited from the text only with difficulty, and appears to be based on a positivistic view of the nature of truth. On page 2 Mr Lucas acknowledges *as a definition* that what is true is what is worthy of [full] belief,[1] a result presumably of the same sort of linguistic analysis that also came up with 'modal' probabilities. Statements of probability possess, though not to the same extent as ascriptions of truth, what the author calls gerundive force, and incline to legitimate *partial* assent. By an easy identification degrees of probability become degrees of 'partial truth': 'Probabilities come out of the same logical drawer as truth and falsehood, interpolated between them.'[2] As if with an eye on the dissident advocate of the correspondence theory of truth, the author accompanies his remarks on the subject with dark animadversions on that theory—it leads to 'infelicities', though one is not told what these are.

---

[1] Lucas, *op. cit.*, p. 1.

[2] *Ibid.*, p. 13.

[3] *Ibid.*, p. 38.

[4] *Ibid.*, p. 26.

That the continuum should provide a Boolean—as opposed to any other—valuation for his language seems purely *ad hoc*. Why this probability logic should satisfy the classical axioms of two-valued logic does not seem to have occurred to the author. We might be inclined to say, classically, that $p \vee \neg p$, the certain event, receives probability one; but only because it can never be false under the classical valuation of the disjunction. There is no apparent reason (certainly none is vouchsafed) why this same formula should take the value—identically true—in the author's system (p. 69). Ironical is his later dismissal of von Mises's theory: 'It attempts to secure by stipulation what needs to be achieved by honest argument, and stipulative stealing is to be rejected not so much on Russell's moral grounds as because it suffers from the disadvantage of extreme fragility.'[3]

Equally *ad hoc* are the continuity and differentiability conditions imposed on the functions representing conjunction and negation. (The monotonicity condition on the numerical function representing conjunction is vitiated by what must be a slip; it is that

$Prob\ (q) > Prob\ (r) \Rightarrow Prob\ (q\ \&\ s) > Prob\ (r\ \&\ s)$[4] unless s is false. But taking s to be $\neg\ q$ the consequent is bound to be false while the antecedent may, it seems, be true.) To round off, one might rejoin to Mr Lucas that his self-appointed task to 'marry the Boolean algebra of propositions' to the real numbers not only lacks justification but the nuptials are moreover suspect. The inspiration of this section of the book is R. T. Cox's pioneering work (Cox [1961]), which however differs sufficiently not to be vulnerable to these objections.

From the point of view of the mathematical development of the theory, of great importance are what the author, following Russell, calls propositional functions. To these he claims to be able to assign fairly precise probabilities; comparably precise evaluations are made of sentences, in general only insofar as they are obtained from such functions by instantiation and quantification. If, for example, we know nothing of Smith, except that he is English, then the probability we shall ascribe to the proposition 'Smith has blue eyes' is just that which is possessed by the formula 'x has blue eyes' relativised to the subdomain of x's defined by the predicate 'English' (cf. chapter 6). This is also an example of a conditional probability, which in this theory is just the probability assigned to some such relativised formula (the notation used by Mr Lucas for designating such formulas is unnecessarily obscure). In other words, conditional probabilities are probabilities of formulas of the form $\varphi^{(A)}(x)$, where A is the condition, to which quantifiers in φ are relativised (this is not the author's notation, but is for all one can see from the text in chapter 6 equivalent to his).

---

[1] 'When we say that something is true we mean that it is worthy to be believed.'

[2] *Ibid.*, p. 22.

[3] *Ibid.*, p. 100.

[4] *Ibid.*, p. 35.

The multiplication theorem and the finite additivity property,[1] the former with an apparently gratuitous derivation on pages 63–4, are easily obtained, with the help of the assumptions noted above, for the probability-bearing propositional functions. From now on, these will, following standard usage, be called open formulas. One of the more curious aspects of this theory is the great emphasis placed on these, and in particular their being assigned values at all. To ascribe truth values to them is certainly unorthodox, as they are usually regarded as only formal entities, convenient building blocks, of which the sentences, traditionally the sole vehicles of content, are constructed. Yet, 'probabilities may properly be assigned to propositional functions, because classical truth values can'.[2] The reason given is that

$$\vdash Fx \quad \text{iff} \quad \vdash \forall x Fx$$

in the usual systems, and 'so to assert Fx is tantamount to asserting $\forall x Fx$.'[3] But this is a perilous argument, for while $\neg Fx$ is the contradictory of Fx, $\forall x \neg Fx$ is not that of $\forall x Fx$, and classically $\neg$ transforms truth into falsity and vice versa. But it is not in general the case that $\forall x \neg Fx$ is true or $\forall x Fx$ is true. The author is saved from inconsistency only by dealing in a many-valued logic, that denies the law of excluded middle (while proclaiming its formal analogue $A \vee \neg A$). However, one scents danger, a scent intensified in noting that only some of the open formulas are in the domain of the probability function (though it is closed under truth functions), the reasons for which will become apparent later. The remainder of this review will fall roughly into two parts, the first pointing out some deleterious consequences of taking probability theory to be a many-valued logic in which open formulas are assigned values; in the second, the primary interpretation as degrees of rational assent will be considered, relative to Bernoulli's theorem and the topic of Bernoulli trials generally.

Now the open formulas define subsets of the domain of variation of the variable (if they contain one variable free); so the algebra of equivalence classes of open formulas is isomorphic to the field of sets they define. Now, the approach of Kolmogorov (Kolmogorov [1933]) is condemned, partly because of alleged shortcomings in the algebra of events; yet in the author's theory, typically in those situations where there is any precise probabilistic reasoning, the prime bearers of probabilities are these open formulas, and the two important axioms are proved for those open formulas. Why then does the author explicitly abjure the word 'event' and emphasise this aspect of Kolmogorov's system in his censure? One has to go all too frequently on clues in this book, in default often of detailed information; but there seem to be two possible reasons. One may be (cf. p. 47) that in Kolmogorov's semantics for his system the space E is regarded as the set of possible 'elementary' outcomes of some trial of the experimental device $\mathfrak{S}$; in other words, the space is a space of, as it were, atomic properties, and the events are properties (in fact sets) of these elementary properties. In particular the space is not one, by intention, of individuals in the ontological sense and consequently no method exists of distinguishing numerically distinct instances of the same property; in the author's theory the mere writing down of the open formula indicates the property and the possible individual of which it is predicted. The other possible reason is that Kolmogorov's fields of sets can include all measurable subsets of the real line for example (they are anyway typically σ-fields). The vernacular of which the author appears so fond has not the resources to define all these 'events'; neither for that matter has first order logic, as he points out. The second is a more plausible explanation perhaps why the author, in a laconic aside, finds the set algebra of Kolmogorovian events 'artificial'. Also of course there is the fact that the author is ostensibly presenting a theory of truth values which are assigned more conveniently to open formulas which at a pinch can be looked upon as degenerate propositions than to events *tout court* which cannot.

---

[1] The author's views on finite *vs.* complete additivity are nowhere explicit; only finite additivity is demonstrated. As a matter of fact, though in standard formalisations only finite disjunctions (and conjunctions) are legitimate, there is a non-problematic extension to the countable case (see Scott [1965]) so there is no immediate formal restriction to finite additivity.

[2] *Ibid.*, p. 47.

[3] *Ibid.*, p. 47.

It was remarked above that not all the open formulas are assigned probabilities. Only those formulas are, for which any sequence of individuals in a suitably specified domain of the individual variable, when examined for the presence or absence of the relevant property, forms a sequence of Bernouilli trials, *i.e.* when each individual has a constant probability of possessing the property and whether it does or does not is independent probabilistically of the other individuals possessing or failing to possess the property. The classic example is 'falling heads' with respect to the set of tosses of a coin in uniform conditions. It is in these circumstances natural to ascribe the probability to the open formula itself, for the probability of an individual in the domain specified possessing the property clearly does not depend on the particular individual chosen. As the whole of the author's theory of what might, by a pardonable ellipsis, be termed physical probabilities, is based on the theory of Bernouilli trials, the way probabilities are to be assigned to open formulas will bear further examination.

Within this theory the probability of a singular proposition is that of the formulas it instantiates, where such a probability-bearing formula exists (it may well not, quite clearly). But subtleties creep in, some of which are dealt with ably enough by the theory, while others appear to be destructive of it as a theory of truth in any, even the 'ordinary' sense, and indicate genuine incoherence.

First, consider the case of singular propositions that do instantiate formulas assigned probabilities, yet describe a past occurrence well-known in its details; *e.g.* 'the last throw of the die yielded a 4'. Clearly, for those with the relevant knowledge, this is worthy of complete belief or complete disbelief, according as the last throw did or did not yield a 4, and so should be assigned a probability of 0 or 1. Yet it is an instance of the formula '*yielded a* $4(x)$' which may have a probability of, say, 1/6. The solution is that we should really consider the singular proposition as instantiating the formula relativised to the subdomain of throws defined by the complete specification; and this will be a subset of the set of throws yielding a 4, or its complement. In other words we should, as far as one can gather from the relevant section of the book,[1] assess the probability of the singular proposition as that of '*yielding a* $4(x)$' conditional upon itself, if the last throw did in fact yield a 4; and this will be 1.

---

[1] *Ibid.*, chapter VI.

The example demonstrates what the author as a matter of fact points out—that probability statements in his theory are referentially opaque (he uses this very phrase of Quine's—see p. 107). This is made quite clear if we use a very similar example as in the preceding paragraph, but let x now range over possible numerical outcomes.

Let $Prob\ (x = 4) = 1/6$

Now the author's theory does, while the standard mathematical theory based on Kolmogorov does not, sanction the substitution of a (previously unspecified) constant a for the variable x, where a denotes the outcome (as yet undetermined) of the next throw. So we get

$Prob\ (a = 4) = 1/6$

However, from the factual identity

$a = 3$

(if the next throw, as a matter of fact, yields a 3) we should get the absurd conclusion

$Prob\ (3 = 4) = 1/6$

upon substituting 3 for a.

This consideration seems to undermine completely his theory as a theory of (even partial) truth. Truth is the fundamental notion of the theory of *reference*; to allow 'referential opacity' here is tantamount to refusing to call a spade a spade. 'It was a demerit of the frequency theory that it could not deal with singular propositions. It is a merit of my account that it can.'[1] But at what cost.

Moreover, it is clear that one and the same proposition may be an instance of more than one open formula; another example is 'Smith has blue eyes'. Smith being a member of each of the classes Englishmen, Northern Europeans, and the species homo sapiens. As each of the open formulas in question will characteristically have been assigned different probabilities, it follows that the same proposition may legitimately be assigned different probabilities, depending upon the class into which Smith is taken to fall. This the author readily admits, an admission the more puzzling as it further undermines the truth-value interpretation,[2] and entails that probability is *not* a function, at least when defined on propositions, despite the notation $Prob(q)$ which implies that it is. The truth-value interpretation is finally and obviously rendered untenable (again if we accept the standard usage of 'truth') by the following consideration: 'Smith has blue eyes' is assigned, by virtue of its instantiating '*Having blue eyes* (x)' a certain probability—say 0·7. What if as a matter of fact Smith does have blue eyes? Then the proposition is just true, irrespective of the formula it instantiates. One does not, to put it mildly, feel inclined to say 'so much the worse for the correspondence theory'. It is no objection to any of these points to invoke the undisputed fact that 'Tomorrow is Wednesday' is true if uttered on Tuesday and false otherwise to support the (highly disputable) proposition that criteria for what is to count as the same proposition invariably refer to the context in which it is uttered (though the author does not in fact appear to resort to this stratagem). 'Tomorrow' is equivocal; 'Smith', assuming the name to refer, is not. It is curious that a writer with so much respect for the 'plain man' (see p. 99 and below) can propose a theory whose principal characteristic is to be so far removed from plain thought.

---

[1] *Ibid.*, p. 101.

[2] 'And so singular propositions puzzle us, because they seem to be straightforward propositions about straightforward public events, but, when candidates for probability values, are regarded as propositional functions and therefore not about substances at all' (p. 107). So 'Smith has blue eyes' is not about Smith at all. This revelation indicates that the author's theory is concerned with *events* (like Kolmogorov's) and is only nominally a probability logic concerned with evaluating sentences.

This is a theory of rational belief or acceptability, yet in which there occur unknown probabilities of open formulas. Let us concede for the moment that there exist unknown degrees of rational belief and see how they are to be evaluated. The first thing to observe is that we might attempt either a point estimate of the probability parameter p, or an interval estimate. In the former case the author recommends a judicious use of Bayes's theorem, commenting that sometimes we do have a reasonably good idea of the prior distribution and that in most cases as long as enough information is obtained the prior distributions do not contribute very much to the posterior probabilities; we eventually get as the best supported hypothesis that which puts p somewhere near or equal to its maximum-likelihood estimator. Or one may simply take note of symmetries and conclude *a priori* (not necessarily finally) that the possible outcomes of some chance device are equally probable. The main tool of inference, to which the author devotes a considerable portion of his book, is Bernoulli's theorem to obtain an interval estimate of p. This is an open challenge to modern and not so modern critics of Bernoulli, who have censured him for trying to do the same thing with his theorem, and it is worth examining to see whether any further light can be shed on this rather murky topic. (It is the reviewer's contention that the author does not succeed in justifying the use he wishes to make of this theorem.)

Before doing this it is as well to reassure the reader that the author's theory is intended to extend further than the simple and elementary topic of Bernoulli trials and the binomial distribution (as a matter of fact he confines himself to the normal when on the subject of continuous distributions, and here it is only dwelt on because of its relation, expressed in de Moivre's theorem, to the binomial, and therefore, to quote the author, 'We can see why it might well be the distribution of the probability densities of some naturally occurring magnitude'); though in view of the necessity of assigning probabilities to the open formulas it is of central importance in this theory.

Continuing with the exposition, let p be the probability of an open formula. Bernoulli's theorem is equivalent to the following statement:

$$\forall \varepsilon, \delta > 0 \; \exists N \; \forall n > N \qquad \mathit{Prob}\left(\left| \frac{S_n}{n} - p \right| \geqslant \varepsilon\right) < \delta$$

where S_n is the sum of the independently distributed random variables x_i, $i = 1, \ldots, n$, where each x_i takes the values 0, 1 with constant probability p, $1-p$. Now '$\left| \frac{S_n}{n} - p \right| > \varepsilon$' is itself an open formula $\varphi_{\varepsilon n}(S_n)$ which defines a new Bernoulli trial.

Clearly one cannot infer that for sufficiently large n

$$\mathit{Prob}\,(p \notin [a_n - \varepsilon, \mathrm{a}_n + \varepsilon]) < \delta$$

where a_n is the observed value of $\frac{S_n}{n}$; for $\frac{S_n}{n}$ is a random variable and a_n is not. What this amounts to saying in the terms of this theory, is that $\varphi_{\varepsilon n}(S_n)$ cannot be instantiated with a_n, for reasons apparent in the previous discussion of the single case. So this expedient is not available. Yet the author claims that von Mises's definition of probability is 'otiose' (in the guise of 'randomness' independence is a presupposition of that theory) since Bernoulli's theorem does effectively 'link probabilities to frequencies.' At first blush his argument is curious, consisting in the observation that 1 and 0 are identified with truth and falsity as part of the theory, while in the semantics we cannot be so precise and therefore identify truth and falsity (in the author's sense of these words, of course) with neighbourhoods of 1 and 0.[1] That seems to be that. But the semantics of the theory are already specified partially in setting up the correspondence between 0 and falsity and 1 and truth. To discuss the semantics of this is then absurd. The suspicion of legerdemain is dispelled (partly) when it becomes clearer (though this section is steeped in obscurity) that the author's argument is actually as follows. Suppose we are testing a hypothesis (about the probability parameter p, say). Every non-zero probability is that of a possible event. So in order to confer content upon the hypothesis, there must be agreement upon some small positive number, such that events occurring with smaller probabilities will be taken to indicate that the hypothesis has been refuted: 'an explanation in terms of a coincidence the probability of which was less than this limit should be acknowledged to be no explanation at all.'[2] The size of the critical region will depend upon the disputants and the case in question. How does Bernoulli's theorem come into it? It asserts that for any disputants and for any Bernoulli sequence, there will come a point after sufficiently many trials when there will be agreement that the probability parameter is approximately equal to the observed relative frequency of successes. For let us suppose that the disputants fix upon a positive number δ_0 such that any event with probability $<\delta_0$ is ruled to be practically impossible. Then it is possible to find an N such that for all $n>N$

$$Prob\left(p \notin \left[\frac{S_n}{n} - \varepsilon, \frac{S_n}{n} + \varepsilon\right]\right) < \delta_0$$

for given $\varepsilon > 0$ as small as desired. The consensus will be that

$$p \in [a_n - \varepsilon, a_n + \varepsilon]$$

for on any hypothesis that set p outside this interval, the outcome a_n would have a critically small probability. This is only a slightly less categorical paraphrase of de Moivre's argument quoted in this book: 'For let the Law be expressed not by the ratio P:Q but by some other, as R:S; then would the ratio of events converge to this last, not to the former: which contradicts our hypothesis' (de Moivre [1718], pp. 251–2).

So: Bernoulli's theorem does not by itself provide the estimate: essentially the procedure is a type of significance test, eliminating a class of hypotheses, with Bernoulli's theorem guaranteeing that, whatever the chosen critical probability, enough trials will confine the parameter within an arbitrarily small interval, with that probability. Hence the link between probability and frequency.

---

[1] *Ibid.*, p. 185.

[2] *Ibid.*, p. 93.

It is as well to pause here and take stock. What are now under discussion are *tests* of hypotheses about probability parameters, and these tests are empirical tests. Indeed, the author frequently mentions 'unbiassed coins'.[1] But probability is not an empirical characteristic, as we saw, for him—according to Mr Lucas, one is not, in saying that $p = \frac{1}{2}$, making a physical prediction about a propensity of a coin and its flipping mechanism, nor a prediction about frequencies; one is not making an empirical assertion at all; indeed, there is a high seriousness adopted towards commission of the naturalistic fallacy.[2] How then can experience refute what we say? It is true, in normal circumstances, that in saying of some empirical sentence σ, that is true, we are making a testable claim; but that is only because 'true' is construed à la correspondence theory. Even had the author showed, which is impossible, that the weak law of large numbers (the usual formulation of the strong demanding complete additivity and hence being underivable here) could deliver zero probabilities, this would still not imply anything factual, even though falsity is identified with probability 0, for his theory of truth is apparently not that of a relation between a language and some interpretative structure, nor does it imply such to exist. His position derives any plausibility it possesses from equivocating, between his own, and the usual semantic theory of truth, and indeed between his own, and an empirical characterisation of probabilities, which leads him to this talk of non-empirical probabilistic conjectures being possible *explanations* of observed frequencies. His position does not even have the purely arbitrary strength of instrumentalism, for he does give a meaning to probability statements (how else could he justify the axioms?) and claims that rejection *rules* based upon observed frequencies are gratuitous. Nor does he adopt the more tenable position of Jeffreys who has a not altogether dissimilar view of the meaning of probability statements, and embrace out and out Bayesianism, for which experience results in changes in *conditional* probabilities, and there is no conjecture or refutation. These observations seem to the reviewer to vitiate the author's view of Bernoulli's theorem and in particular his claims against the frequentists. How could Bernoulli's thorem provide frequencies out of thin air? Conclusions about frequencies follow only from assumptions about them or from contradictions. The frequentists, it must be admitted, cannot derive frequencies in finite samples from their hypotheses, but at least they face up to the realities of logic. The most that is forthcoming from Bernouilli's theorem for the author, is that one is justified to an extent greater than $1-\delta$, $\delta<0$ in betting against $p \notin [S_n-\varepsilon, S_n+\varepsilon]$—not, by the way, that here $1-\delta$ is itself necessarily any sort of betting quotient—and nothing more. In particular, it cannot be inferred that the occult p lies in any given interval obtained experimentally. Since p denotes no physical characteristic, it is difficult to see what hinges on such a bet, anyway. It might be said, however, that either p is in the interval or it is not; and this is a purely objective matter, so that the problem is as important as any scientific conjecture; it is also a statistical matter, and so susceptible to the usual statistical tests or estimation procedures. The simple answer to this is that the principal part of the justification of these methods is usually couched in terms of frequencies *viz*. Fisher's defence of significance tests: 'either an extremely rare event has occurred, or the hypothesis is false' (Fisher [1956]). To repeat, there is nothing about frequencies, or anything empirical, in the author's understanding of the statement '$p = p_o$'; no experience therefore can in any way impinge on it. Desideratum (i) is not satisfied.

---

[1] *Ibid.*, p. 88.

[2] *Ibid.*, p. 28.

One can also and ought, to dispute the premise, that it is an objective matter of fact that p is in a given interval, or not, in the author's theory, for a closer scrutiny casts doubt upon whether p exists at all. By and large one may say that any theory of the semantics of probability that concedes the widespread existence of Bernoulli trials, and of independent events generally, has passed the first and most necessary hurdle. It is in fact extremely doubtful that the author's theory can do this, and that it appears able to depends on the equivocation already remarked. Consider a sequence of tosses of a coin under uniform conditions. It is easy to say that here there are almost certainly independence and constant probabilities; this is the paradigm of a Bernoulli sequence. But if we attend to the author's interpretation of p, then this ceases to be so certain. For how can the statement 'heads occurs at the nth trial' be independent in terms of its degree of rational assent, of statements describing the outcomes of the first $n-1$ trials? Certainly not simply because there is no causal connection between outcomes of trials, for it does not by any means follow from this that a posterior assessment of heads must be the same as the prior; to find out the chance of heads it is after all only rational to observe actual frequencies; however, the chance of heads must be a physical characteristic of the set up. Only by appealing to such on which to base rational degrees of belief can the latter have the independence property. This objection was first voiced (though not to this theory) by de Finetti thirty four years ago (de Finetti [1937]); he too was unwilling to countenance physical probabilities and was forced to abandon the notion of independence and rely on the weaker exchangeability property.

It may be noticed how a failure to grasp the equivocation leads one to concur without thought in the existence of Bernoulli trials within this theory and to identify them with among other things the standard coin-tossing procedures. And in this we are aided by the author's rather eccentric use of the word 'truth'. For if we are sensitive to the connotation of the word 'true', of correspondence with reality, then certainly it is easy to grant that truth-assessments may well be independent of what has gone before (it is a trite enough remark that enumerative induction is invalid) without even having any clear idea what intermediate truth values are. The upshot of all this is, then, that even if we accept the author's work as a theory of truth values (which independently the reviewer has urged we should not), and if we grant the existence of Bernoulli trials on the basis, perhaps, of deep-lying ideas of truth, it follows that the central plank of the author's account—the identification of degrees of probability and degrees of probability and degrees of rational belief—will not bear any weight.

This is enough exposition and criticism of the author's own positive views to give the reader some idea of the substance of the book. Now I shall discuss the author's strictures on the subjective theory and the frequency theory of von Mises.

The subjective theory is skimpily treated: one of the author's charges, in one of the useful appendices at the end of the book, is that 'it cannot account for reasonable, intersubjective assessment of probabilities'; but this is to ignore entirely the boast of its subscribers, and notably of one of its founders (de Finetti) that this is precisely what, in many cases, it can do. Prior probabilities with respect to exchangeable events, are dominated by sufficient evidence just as prior probabilities with respect to independent events are dominated, via Bayes's theorem (as the author shows with admirably clarity), by sufficient evidence; the result in both cases tends to be the same: one of the hypotheses will eventually come to be overwhelmingly supported. The philosophical importance of these formulas of the subjective theory may be disputed, but they should be mentioned, as together with the Ramsey-de Finetti theorem (also not mentioned) they constitute the substance of the subjectivists' claim to have solved 'the problem of induction'. There are two other errors; one is that the subjectivistic theory 'assumes that people who bet are rational', and the other, adduced as a merit of the theory, no less, is that 'it is based on the betting propensities of the British people'. Either these two allegations are amusing howlers or the author has his tongue in his cheek. The only other criticism of subjectivism in this appendix devoted to other theories, is rather mystifying; it is that this theory 'erects the theory of probability on an extremely slender basis'. Presumably this is an allusion to the fact that it denies the existence of stochastically independent events; part of the burden of this review is that in this it is no different from the author's own theory.

Finally, to the theory for which the author reserves his biggest guns—the frequency theory of von Mises. This is the only frequency theory that the author considers (though many of the shortcomings of this theory are not attributable to later versions—in particular the inability to deal satisfactorily with the single case. This is not to say that these subsequent versions, like Popper's, for example, may not themselves possess demerits; but they should be mentioned). First it is as well to dispose of some of the author's more trivial objections: for example, that this is not a 'plain man's' theory, as if *Probability, Statistics and Truth* were a notorious compendium of bizarre and exotic fantasies. The author's own theory is hardly a plain man's theory either, for that matter. As for the claim that the frequency definition is otiose, that has been already and adversely dealt with: not even a theory able to deliver the strong law of large numbers in its strongest form can write off without prior stipulation zero probabilities as practical impossibilities. The most serious criticism is one that is not original to the author: it is that the definition proposed by von Mises is actually inconsistent with the binominal formula. It may seem impossible to urge this criticism simultaneously with the preceding one. Be that as it may, it, like the other one, fails to hold water. In answer to the question why the definition should seem to conflict with the binominal formula one has only to note that in the case where the probability of success relative to some collective is $p \neq 0$ or 1 then there is a finite probability that for any sufficiently small real number ε and for any positive integer N there is an integer n such that the absolute difference between the relative frequency of success in $N+n$ trials and the probability p exceeds ε: one simply has to suppose n consecutive successes occur after the Nth trial, and the probability of this is $p^n > 0$. Consequently there is a finite probability of the collective diverging from p, or so it seems. However the apparent paradox is easily dissolved.[1] The first step consists in noting that when discussing the properties of the random variable S_n the sample space implicitly referred to is a set of sequences of 0 and 1. The above 'objection' is equivalent to the assertion that if we choose an arbitrary value of N then it is possible to find sequences in this set which have the relevant number n of consecutive successes after the Nth place (in other words the convergence is not uniform over this set), such that if the original collective is partitioned into successive subsequences each of $N+n$ elements then the limiting relative frequency of these subsequences in which the last n terms are successes will be p^n. The theory is coherent.

---

[1] The following argument first, to my knowledge, appeared in W. C. Kneale's [1949].

This book may be hard to adjust to, let alone follow, and this will not always be because of the frequent obscurities in the author's presentation of his case, but because a necessary acquaintance with the deliberative processes typical of adherents to Oxonian philosophy is lacking in the reader. For example, it is not an obvious consequence of the truism that probability is a 'guarded guide' to life, that the term does not denote some empirical characteristic. If in fact probabilities were propensities to produce, in a determinate manner some event then they would clearly support an equivalent betting quotient, and hence the adjective 'probable' (indicating a probability greater than some accepted non-commital quantity) would have 'gerundive force'. It thus does not follow that possession of that virtue necessitates a non-naturalistic definition of the term. G. E. Moore discovered, in attempts to provide a naturalistic definition of the word 'good', the naturalistic fallacy, and inferred that goodness is non-natural (as well as unnatural). He may have been right; followers of Wittgenstein and Austin conceded that it was non-natural, but seemed to deny that it denoted anything at all (Moore thought it did). Here is the extension of that doctrine to probability theory, issuing in this notion of modal probability, and vitiated, it seems to me, by the difficulties ensuing. It is likely that nothing worthwhile (outside, possibly, moral philosophy) will ever come out of this general approach to philosophical problems; principally because it isolates a Platonic 'meaning' as the focus of philosophical scrutiny, at the expense all too often of the intention of the scientific innovator who is usually more concerned with the class of phenomena he wishes his terms to delimit, and for whom a sensitivity to nuance is not in general permitted to dictate his methods of theory construction, even where he may use, while making more precise, terms that in the vernacular have associations in conflict with the meanings he chooses to adopt. 'Words mean what I choose them to mean; it is a question of who is to be master, that's all.' Thus White, exemplifying in his [1972] a type of linguistic philosophy, can write 'In trying to discover the exact meaning of the propensity theory of probability, I have frequently alluded to the contrast between *being liable* and *being likely*', and go on to suggest that the propensity theory is fallacious since it ostensibly conflates these two notions (and also habits and predispositions). Whether it does or whether it does not does not seem a very important issue beside the question of whether certain adequacy criteria postulated by its advocates, like those outlined at the beginning of this review, are satisfied. One can imagine Frege's reaction to kindred censure over his definition of the number one. It would have been brief but pithy, one feels.

Mr Lucas's book, to repeat, is one that does address itself to real issues, and wrestles with them; it is besides a work of great charm. But its limitations are real, deep and ultimately philosophical.

COLIN HOWSON
London School of Economics and Political Science

REFERENCES

Cox, R. T. [1961]: *The Algebra of Probable Inference.*
De Finetti, B. [1937]: 'La Prévision: ses Lois Logiques, ses Sources Subjectives', *Annales de l'Institut Henri Poincaré*, **1**, pp. 1–68. Reprinted in English in H. E. Kyburg and H. Smokler (*ed.*): *Studies in Subjective Probability*, 1964, pp. 93–158.
De Moivre, A. [1718]: *Doctrine of Chances* (Page reference refers to the third edition).
Fisher, R. A. [1956]: *Statistical Theory and Scientific Inference.*
Kendall, M. and Stuart, A. [1951]: *The Advanced Theory of Statistics.*
Kneale, W. C. [1949]: *Probability and Induction.*
Kolmogorov, A. N. [1933]: *Die Grundbegriffe der Wahrscheinlichkeitsrechnung.*
Popper, K. R. [1959]: 'The Propensity Interpretation of Probability', *British Journal for the Philosophy of Science*, **10**, pp. 25–42.
Scott, D. [1965]: 'Logic with Denumerably Long Formulas and Finite Strings of Quantifiers', in J. W. Addison, L. Henkin, and A. Tarski (*eds.*): *The Theory of Models*, pp. 329–41.
Toulmin, S. [1958]: *The Uses of Argument.*
White, A. R. [1972]: 'The Propensity Theory of Probability', *British Journal for the Philosophy of Science*, **33**, pp. 35–43.

THE SIGNIFICANCE TEST CONTROVERSY*

1 *Introduction.*
2 *Basic Concepts of Significance Testing.*
3 *The Logic of Testing and the Methodology of Research.*
4 *Experimental versus Non-Experimental Research.*
5 *Testing Theories and Causal Hypotheses.*
6 *Problems in the Logic of Testing.*
7 *Alternatives to Significance Testing.*
8 *Conclusion.*

1 *Introduction.*

Since the end of World War II the place of mathematics in psychology and the social sciences has been firmly established. This is true with respect both to the content of hypotheses and theories and to the methodology employed. The methodological role of mathematics has been most evident in the use of statistical methods, particularly the use of so-called tests of significance. Indeed, in recent decades the demonstration of a 'significant difference' has sometimes been taken as virtually a necessary condition for the publication of empirical studies in these sciences. But the hegemony of the significance testing methodology has not been complete. Articles critical of various aspects of this methodology have found their way into the journals and some are now being reprinted in numerous books of readings in statistics and social science methodology. The most comprehensive of these latter, at least as regards significance testing, is *The Significance Test Controversy*, edited by two sociologists, Denton E. Morrison and Ramon E. Henkel.[1]

---

* This work was completed during the author's tenure as associate research scientist at the Courant Institute of Mathematical Sciences, New York University, and has been supported in part by the National Science Foundation under grants GS 2525 and GS 28628. Professor Allan Birnbaum's helpful criticism and hours of patient discussion on these and related problems are hereby gratefully acknowledged.

Given that the volume contains mainly criticism and only vague hints of possible alternatives, one must hope that its intended readers will gain sophistication and not lapse into methodological anarchy out of despair or confusion. But whatever the effect on its intended audience, this volume contains a wealth of material and references regarding issues which should concern any serious student of probability, induction and scientific method. The time is long past when philosophers of science, in particular, could concern themselves solely with the ultimate justifiability of inductive methods and other puzzles largely of their own making, *e.g.*, the paradoxes of confirmation. Our goal must be to develop comprehensive, unified accounts of inductive reasoning which can be applied to actual research in the sciences. To do this one must be familiar with the kinds of methodological problems actually faced by investigators in many fields since it is these problems which any adequate comprehensive account of inductive methods must help resolve. The great virtue of *The Significance Test Controversy* for philosophers, therefore, is that it provides easy access to discussion by sociologists and psychologists of a range of methodological problems as they see them in their respective fields. Hopefully this availability will enrich the philosophical literature and eventually pay dividends in the form of methodological clarity in sociology, psychology and in other fields as well.

The book itself is divided into five parts. The first part, intended to give the 'critical historical context', consists of three selections from Hogben [1957]. Even the editors admit, however, that these selections are not entirely appropriate —not because they are bad, but because on all levels, *i.e.* historical, scientific and methodological, they have much wider scope and provide a deeper analysis of statistical theory and practice than any of the selections that follow.[2] The second and third parts deal with 'the controversy in sociology' (19 selections) and 'criticism by psychologists' (5 selections) respectively. This division is reasonable since the two groups have focused on different problems and there is no recorded interaction between them. The fourth part, 'criticism from other quarters', contains three papers which do not fit elsewhere. The last section, 'epilogue', contains a brief and rather academic, though hardly neutral, summary of the issues by the editors. This is followed by ten pages of references and an index.

2 *Basic Concepts of Significance Testing.*

Morrison and Henkel assume their audience to be familiar with standard statistical tests used by sociologists, *e.g.*, the t-test. For the uninitiated, a brief examination of any standard undergraduate text in statistics will open the door to most of the issues raised in these papers.[1] For the purposes of this review, the following sketch will be sufficient.

---

[1] Morrison, Denton, E. and Henkel, Ramon, E. (*eds.*) [1970]: *The Significance Test Controversy: A Reader: Methodological Perspectives.* Chicago. Aldine Publishing Company. Pp. xviii+335. Other volumes including similar and related material are: E. F. Heerman and L. A. Braskamp (*eds.*): *Readings in Statistics for the Behavioural Sciences*, 1970; B. Lieberman (*ed.*): *Contemporary Problems in Statistics*, 1971; J. A. Sleger (*ed.*): *Readings in Statistics for the Behavioural Scientist*, 1971; E. R. Tufte (*ed.*): *The Quantitative Analysis of Social Problems*, 1970.

[2] A sociologist of science might find it interesting to investigate the apparent non-reception given to Hogben's book by the statistical establishment.

A statistical hypothesis describes a distribution of probabilities over the possible outcomes of a chance process. One of the most familiar kinds of chance process is the random selection of a sample from a population. Taking an example discussed by Selvin and others, one might sample the residents of a city and record their mean level of 'political interest' determined by some presumed reliable indicator. Now suppose we are interested in whether urban and rural residents have different levels of political interest. We might then also sample the residents of a nearby 'rural' area and take the difference in means of the two samples as the outcome of the dual sampling process. Intuitively one would be inclined to reject the 'null hypothesis' that there is no difference in political interest if the observed difference in mean levels is large. The operative principle here seems to be that a statistical hypothesis is to be rejected (regarded as false) whenever the observed outcome of the relevant chance process would be very improbable if that hypothesis were true. Something like this principle has been employed by investigators for over 250 years.[2]

In the 1920's, R. A. Fisher developed mathematical techniques for applying this basic idea precisely in a variety of experimental contexts, primarily in genetics and agricultural experimentation.[3] Fisher taught that one must consider all possible observable outcomes of a sampling experiment and then designate a subset of these as indicating that the test hypothesis is to be rejected. The probability of obtaining a result in this rejection set if the test hypothesis is true is called the 'significance level' of the test. The significance level thus designates the probability that the test will lead to the 'mistake' (later called Type One Error) of rejecting a null hypothesis when it is true. In his writings and practice, Fisher generally recommended using significance levels of at least ·05. Following this practice one could expect to reject a true hypothesis at most five times in a hundred.

In the late 1920's and early 1930's, Neyman and Pearson developed a more systematic theory of statistical hypothesis testing which fills apparent gaps in the Fisherian approach. Continuing the previous example, suppose the observed difference in mean political interest were 'not significant' at the 0·05 level, *i.e.* had a probability greater than ·05 of occurring by chance sampling variations alone. Do we conclude that there is no difference, *i.e.*, do we 'accept' the null hypothesis, or do we simply note that we failed to reject it? Fisher's pronouncements and practice are ambiguous on this point. Neyman and Pearson wish to accept the hypothesis, but this requires consideration of a second type of possible error, *i.e.*, accepting a false hypothesis. The probability of making a Type II error, however, is only definable relative to an alternative statistical hypothesis.

---

[1] A highly recommended text would be Hays [1963].

[2] In the early 1700's J. Arbuthnot observed that there had been more male than female births registered in London for 82 consecutive years. He used this data to conclude that the probability of a male birth is not one-half. Associating chance with equal probabilities, he concluded, with apparent seriousness, that divine providence must be operative in the birth of children. The argument is recorded in Arbuthnot [1710], and has been discussed recently in Hacking [1965] and Birnbaum [1967]. Laplace also made liberal use of simple statistical tests in examples discussed in his well-known [1814].

[3] Fisher's ideas reached the scientific community primarily through two books (Fisher [1925], [1935]). Both volumes went through numerous editions during Fisher's lifetime.

Given a specific alternative, one can calculate the probability of both mistaken rejection and mistaken acceptance relative to any possible outcome of the experiment. Neyman and Pearson showed how best to choose the rejection region of possible outcomes so that the 'power' of the test, *i.e.*, the probability of rejecting the null hypothesis if the alternative is true, is optimised for any fixed significance level. Thus what differentiates Neyman-Pearson tests from simple Fisherian tests is an explicit concern with alternative statistical hypotheses and with the power of the test relative to these alternatives.[1]

The above sketch exaggerates the differences between Fisherian significance testing and Neyman-Pearson hypothesis testing. Moreover, it does not mention possible gaps and difficulties in the Neyman-Pearson theory. Nevertheless, one should be aware that there are two approaches because many current criticisms of statistical testing apply only to the Fisherian variety and could even be taken as support for the Neyman-Pearson approach. To take an example, one argument which occurs frequently in both the sociological (*e.g.*, Kish) and psychological (*e.g.*, Bakan) literature goes as follows: on general empirical grounds we know that there is almost always *some* small difference between any two parameters. Thus we know when we start that the null hypothesis is false. If our test fails to reject it, that only tells us we did not take a sample big enough to detect the difference, not that there is none. So why bother testing in the first place?

Fortunately even Fisherians have had the good sense to ignore this argument, though they may have some difficulty saying why it should be ignored. The Neyman-Pearson account, on the other hand, allows us to make the obvious response, namely, that we are not concerned with just any difference, but only those large enough to be distinguished by our current theories and experimental techniques. Such differences are reflected in the low power we require for alternative hypotheses near the null hypothesis. Any test should of course have relatively high power against importantly divergent alternatives. These power requirements together with the chosen significance level determine the appropriate sample size for the test. Thus, on such an understanding of the Neyman-Pearson account, we are effectively testing whether the true value is near the null value, or importantly far from it, and this is surely worth doing.[2]

3 *The Logic of Testing and the Methodology of Research.*

In reviewing the significance test controversy it is helpful to distinguish several kinds of criticisms of such tests. Most interesting to philosophers and theoretical statisticians are arguments that hypothesis testing misrepresents the form of inference appropriate to scientific research. Second are arguments that tests do not provide grounds for certain types of conclusions, *e.g.* causal hypotheses or theories. Third are general arguments that tests are inapplicable to important types of research, *e.g.* non-experimental. Fourth are arguments that tests are frequently misinterpreted as indicating something they obviously could never indicate, *e.g.* substantive importance. Finally there are arguments that tests are simply misapplied, *e.g.* to cases in which necessary empirical assumptions are known to be false. This latter type of objection is of no philosophical interest. The fourth type of objection raises questions that are, or should be, of interest to philosophers of science. These questions merit a few remarks.

---

[1] The original Neyman–Pearson papers (1928-38) have been reprinted as Neyman and Pearson [1967]. Cf. also Neyman [1950]. For a most instructive comparison of Fisher's and Neyman's approach to testing, read Fisher's discussion of the 'lady tasting tea' in Fisher [1935], Ch. II, and then read Neyman's commentary in his [1950], Ch. V. For an elementary exposition of testing from Neyman's point of view cf. Hodges and Lehman [1964].

[2] For a technical elaboration of this position, cf. Hodges and Lehman [1954].

A number of authors, *e.g.* Kendall, Selvin, Gold, complain that too many researchers seem to confuse statistical significance (in the technical sense) with substantive importance since they obviously worry more whether their results are statistically significant than whether they contribute to the theoretical development of the field. In context such remarks seem intended to impugn the use of tests as rules of inductive validation. But clearly they can have no such implication. At best a test can only tell us which hypotheses we may with low risk take as true (or false); it cannot tell us which are interesting. An analogy may illuminate this point. Imagine a mathematician who complained that too many people were deducing and publishing theorems of little or no importance for the development of mathematics as a whole. Most would regard this as an unfortunate fact about mathematicians and the organisation of the discipline. But who would take this as a criticism of the rules of deductive inference?

If statistical significance provides no criteria for substantive importance, what does, and how might one best seek results that are both statistically and substantively significant? These questions are often raised, *e.g.* by Gold and by Morrison and Henkel, but not developed in these papers. Morover they differ from other questions raised in that they bear more on the methodology of research than on what a philosopher would call the logic of testing. This is not to say, as some philosophers would, that strategies for pursuing substantive significance are at best part of the subject matter of psychology, not its methodology. Indeed, useful books are currently being written by statisticians (*e.g.* John Tukey) on 'exploratory data analysis'. Nevertheless, the main issues in the significance test controversy concern aspects of the logic of testing, and it is these that are discussed below.

4 *Experimental versus Non-Experimental Research.*

Most of the well-known objections to significance testing have appeared in the sociological literature, but 'the controversy in sociology' has been dominated by the fact that sociology is almost totally a non-experimental science. In experimental psychology, for example, one typically selects two groups at random from a population and applies a specified treatment to members of one group while keeping the other as a 'control'. If random selection is achieved, then the probability of any difference in the measured dependent variable can be calculated relative to the hypothesis of chance alone, *i.e.* supposing the treatment has no effect. Thus the significance level (and the power function too) of the test are well-defined and knowable.

In sociology, on the other hand, one may sample one or two populations, *e.g.* urban and rural, but the populations themselves will have been self-selected through a process about which we know very little. In some cases one may even examine the whole population of interest, *e.g.* the students in a medical school (Kendall), so that there is no randomisation of any kind in the actual investigation. What can be the point of calculating a significance level when the condition of random sampling or random assignment is known to be lacking? The editors, following Selvin, Kendall and others, insist that the use of a significance test requires a random process in the generation of data and that this is lacking in most sociological investigations. Thus, they conclude, the use of tests in sociology merely gives the illusion of scientific validity and diverts attention from the main task which is developing deductive theories.

The contrary view is represented, *e.g.* by McGinnis, Gold and Winch. They hold that sociologists are implicitly concerned with stochastic processes even in investigations involving no explicit sampling. Suppose we sample from two self-selected populations, *e.g.* urban and rural, and find a significant difference in mean political interest. We are certainly justified in concluding such a difference exists. There is, of course, little basis for conclusions concerning possible *causal* relations between place of residence and political interest. But contrary to some authors, *e.g.* Selvin, the *statistical* inference is still valid without random assignment and is, moreover, necessary to the further causal inquiry (see Section 5 below). Even if one exhausted the populations in question, there is still a chance process to be the subject of the null hypothesis, namely, the historical process which produced the populations in question. In rejecting the null hypothesis one concludes that this process tends to produce different levels of political interest in urban and rural settings.

Whether or not social phenomena are intrinsically deterministic or indeterministic, it is surely true that relative to our present knowledge, most social configurations can be regarded, for most theoretical and practical purposes, as the result of a stochastic process. Why have serious sociologists been reluctant to acknowledge or make use of this fact? Part of the reason is that they have taken random sampling as the paradigm stochastic process so that in the absence of explicit sampling they are unwilling to talk about probabilities at all. In this they have unfortunately been encouraged by some statisticians, especially Fisherians (see the selection from Hagood's text), who have taught that an existing population should be viewed as a random sample from a hypothetical infinite super-population of possible populations. It is hardly surprising that sociologists have found such talk excessively abstract and obscure. But it is so only because of too close an identification of probabilities with infinite frequencies. If, following statisticians like Cramér, we take probabilities to be more theoretical quantities, something like Popper's 'propensities', then we may legitimately assign probabilities to possible outcomes of a unique process.[1] Once this conceptual barrier is overcome, the only remaining problem is having enough knowledge of social processes to construct a statistical test. But fortunately the knowledge needed for many interesting tests is minimal, and experienced investigators can agree on reasonable assumptions to be employed in relatively 'robust' tests.

5 *Testing Theories and Causal Hypotheses.*

Everyone knows that a statistical hypothesis never logically implies a causal hypothesis, though some empiricist philosophers would like to analyse causal statements into a set of related statistical statements. Assuming that the primary goal of any science is to develop explanatory, causal theories, can tests of merely statistical hypotheses have any important role in scientific inquiry? The critics of significance testing, *e.g.* Selvin, Camilleri, Morrison and Henkel, seem to be saying 'No'. But a negative response assumes that statistical hypotheses do not play an essential role as *evidence* for the confirmation of causal theories. If statistical hypotheses do play this role, then clearly some kind of test for the truth of statistical hypotheses is also essential to scientific inquiry.

---

[1] Cf. Cramér [1946], Ch. 13. Popper's most recent views on probability which are as ambiguous as they are suggestive are in Popper [1967]. For a recent discussion and further references, cf. Giere [1971].

Assuming a satisfactory analysis of probabilistic causality, we can say that a causal theory has statistical consequences regarding many actual (and infinitely many possible) populations and stochastic processes. Some of these consequences are selected for investigation and tested. If the rate of rejected consequences is about equal to the significance level of the individual tests, we have some grounds for concluding that the theory is at least approximately correct. In effect we are here applying a statistical test to the statistical meta-hypothesis that all consequences of the theory are true. And this meta-test takes account of the small probability (equal to the significance level of the individual statistical tests) that a true consequence will be judged false.

That something like the above approach to theory testing is at least implicitly employed in behavioural research is granted even by the opponents of significance testing *e.g.* Meehl, who apparently questions the above account of theory testing as well. Meehl's argument deserves special comment. He begins by noting that any simple null hypothesis is overwhelmingly likely to be false. In experimental contexts this means that every treatment has some effect, however small, on any measurable characteristic. Meehl next observes that theories in psychology (and other 'soft' sciences) typically predict only directional statistical hypotheses, *e.g.* that a difference in means is greater (or less) than zero. This prediction is then checked by subjecting the null hypothesis, *i.e.* no difference, to a significance test. If the null hypothesis is rejected and the difference is in the predicted direction, the theory is confirmed. Now for the sake of argument suppose there is some antecedent probability that the difference is in the predicted direction and that the maximum value of this probability is less than 1.[1] It follows that for any experiment the antecedent probability of the test confirming the theory will be *increased* toward its maximum by our increasing the power or 'precision' of the test, since this increases the probability that *any* difference will be detected. This result, Meehl claims, is paradoxical since in the physical sciences the antecedent probability of corroborating a theory *decreases* toward zero with increasing precision of measurement. It is assumed, with Popper (and the Bayesians), that the degree of corroboration of the theory is inversely proportional to the antecedent probability of the evidence.

One wonders why Meehl thinks the above result is a real paradox. The crucial difference between physics and psychology, as Meehl himself points out, is not that theories are tested differently in these fields. It is that psychological theories have low content, they predict only directional hypotheses, while physical theories have high content, they predict point valued hypotheses. If a psychological theory implied a point hypothesis, this would be the null hypothesis, and the antecedent probability of a confirmation, assuming such exists, would surely decrease with increasing precision. Similarly, if a physical theory yielded only directional consequences it would be tested as psychological theories are tested. The fact that working with low content theories tempts psychologists into methodological sin, *e.g.* introducing *ad hoc* auxiliary hypotheses, is hardly the fault of the way statistical hypotheses or theories are standardly tested. Nor is there any reason to think physicists would be any more virtuous in a similar situation.

---

[1] Meehl gives an ingenious argument to show that with maximum precision the antecedent probability of obtaining a positive result is at least half. But this argument is not essential to the stated paradoxical conclusion. It is enough that there be an antecedent probability and that its maximum value be less than one.

Finally, what are we to do if Meehl is correct in thinking the construction of more precise theories in psychology is premature? The answer, suggested by Lykken, is to derive and independently test numerous directional hypotheses. The antecedent probability of n independent successful tests should be of the order of $(\frac{1}{2})^n$, thus providing high corroboration for the theory with a moderate number of tests. This procedure, of course, assumes that our theories are comprehensive and unified enough to yield sets of 'different' consequences, even if these are only directional hypotheses.

6 *Problems in the Logic of Testing.*

None of the lines of criticism discussed above raises fundamental questions about the internal logic of testing itself, at least not directly. This is characteristic of the significance testing literature. Yet it is not that questions about the logic of tests are never raised; they are just not so much emphasised or developed as questions about the applicability of tests in certain circumstances, *e.g.* non-experimental research, or for certain purposes, *e.g.* confirming causal theories.

One frequent objection is that the choice of ·05 and ·01 as the proper significance levels is arbitrary. This question may not seem so fundamental if what one desires is simply a rationale for using one level rather than another. For while the significance level (and power too) is not determined by the bare mathematics of significance testing, and thus 'arbitrary', the proper level may be determined by other factors in the scientific context. Fisher frequently hinted at such a justification by saying that we can afford to be misled (*i.e.* to reject a true null hypothesis) one time in twenty. This seems to imply that research would stagnate or disintegrate into confusion if our rate of accumulating Type I errors were much higher. Similar considerations apply to the precision or power of tests. Of course such constraints could never justify ·05 over ·06 as the correct level, but the desire for such discrimination when none is possible may be viewed simply as a human failing of researchers and editors. If Fisher is on the right track, one mistake in four would be a disaster, but one in ten perhaps not so horrible.

Neyman and his followers, particularly Wald, came to view hypothesis testing as a special case of decision making under uncertainty.[1] In this general framework one begins with a set of possible actions, a set of possible states of the world, and a utility assignment for each possible outcome, *i.e.* each action-state combination. The optimal action is then a function of the utility assignments. Thus in hypothesis testing the role of significance level and power is taken over by utility assignments to the outcomes of making or not making Type I and Type II errors. Furthermore, on a strict interpretation, to accept a hypothesis is just to perform some definite action.

---

[1] Cf., for example, the introductory chapter to Lehman [1959].

Here it is clear that the problem of rationally choosing a proper significance level leads to a more fundamental problem, namely, can *any* significance level be rationally justified? More generally, should empirically testing a statistical hypothesis be viewed as a decision process at all? This question is raised by Camilleri, Skipper, Morrison and Henkel, Bakan and, most clearly and forcefully, by Rozeboom. The negative argument is that the actions and utilities required by any explicit decision approach are available only in practical contexts, *e.g.* in industrial quality control or business decisions. In the theoretical context of 'pure' research, it is claimed, the needed actions and utilities simply do not exist.

The question whether one can speak of scientific 'conclusions', as Fisher does, without presupposing the whole decision theory framework is not raised in these papers. Nor does anyone mention the possibility, suggested by Fisher's remarks and discussed in limited contexts by several philosophers, *e.g.* Hempel and Levi, of introducing distinctively 'epistemic' or 'scientific' utilities.[1] These issues go considerably beyond the significance test controversy into the deeper waters of theoretical statistics and the foundations of scientific inference. Nevertheless, the discussion of such issues in the significance testing literature is important because it indicates just where the deeper discussions of statistical inference must match up with the concrete research concerns of psychologists and social scientists.

7 *Alternatives to Significance Testing.*

It was no part of the editors purpose systematically to survey alternatives to significance testing, but some alternatives are discussed and are worth mentioning here. A number of the sociologists, Selvin and his associates, Camilleri, Morrison and Henkel, suggest abandoning testing in favour of the development of systematic deductive theories. This is an extreme reaction, though understandable in light of the amount of research apparently designed primarily to show some 'significant difference'. But without relatively precise criteria for judging statistical consequences, how could one ever test a theory against basic survey or experimental data? Or, as Davis puts it, without tests of statistical hypotheses, how does one know what is to be explained? Deductive coherence on the theoretical level is not enough.

One popular suggestion is to use interval estimates in place of tests. But it is not clear that estimation is always appropriate; nor is it clear that estimation avoids the objections to significance testing. If a theory predicts the form of a statistical distribution but leaves a parameter unspecified, interval estimation of the parameter is clearly in order. But if the theory predicts a unique distribution, a test of that point hypothesis seems most appropriate. Of course, one might use an interval estimate like a test, accepting the hypothesis if the predicted value falls in the interval and rejecting it otherwise. But then the difference between testing and estimating is merely verbal since the confidence level and width of the interval are strict analogs of the significance level and power of a test. The same sorts of considerations apply in both cases. Perhaps the admonition to estimate rather than test is best seen as part of a plea to give up the cataloguing of 'significant differences' and to begin building theories of the sort which will have parameters in need of estimation. The significance testing literature reveals many cases of scientists blaming their tools rather than themselves for the slow progress in their fields.

---

[1] Cf. Hempel [1962] and Levi [1967].

Another suggestion, which needs to be worked out in detail, is to report the significance level (and power!) actually achieved, rather than reporting an acceptance or rejection at a standard level. Similarly, one could report the confidence level and width of an interval estimate This suggestion serves the useful function of allowing sociologists to report studies with significance levels lower than the sacred ·05, which, one should remember, was introduced in biological experimentation It would be better, however, to argue directly that higher levels are appropriate in sociology at the present time

The suggestion to report observed levels rather than the acceptance or rejection of hypotheses is really a suggestion to report only the data (in a reduced form) and not to report conclusions It seems to be assumed that each reader will draw his own conclusions. But in this case all the objections to significance testing will still have to be faced by each individual. Moreover, the idea that there is a body of public knowledge going beyond mere reports of experiments is lost. One might, of course, take the view that no one ever accepts or rejects anything (except the data), but this leaves us wondering what one ever does with all the reports of levels and how these might be related to the theories everyone so badly wants.

The above alternatives are really attempts to improve on the significance testing methodology without radically altering one's basic conception of scientific inference. Rozeboom is the only contributor to this volume who explicitly advocates a radically different alternative, namely a Bayesian theory of inference. Indeed, Rozeboom thinks of cognitive inquiry as essentially a process of altering the degree of probability of hypotheses in the light of data. This process involves inverse probabilities and relies heavily on Bayes's theorem. Rozeboom, however, rejects the interpretation of probability relations as merely subjective degrees of belief and holds out for some type of logical relation. In this way he avoids the conceptual ties between subjective probability and decision making so prominent in the views of well-known Bayesians like Ramsey or Savage. It is, however, not clear how Rozeboom would avoid the difficulties with logical probabilities that have led others to embrace some form of subjectivism.

There are, of course, other alternatives not represented in these selections. Some theoretical statisticians, for example, seek to evaluate evidence in terms of likelihood ratios alone while others have sought to revive Fisher's fiducial argument. *The Significance Test Controversy*, however, is only concerned with such alternatives in so far as they bear on the every-day practice of behavioural scientists. Thus far theoretical statisticians outside the Fisher-Neyman-Pearson tradition have been slow to bring their views to bear on such problems.[1] Needless to say, philosophical students of probability and induction have been even slower.

---

[1] The Bayesians, however, have made some efforts in this direction. Cf. Edwards, Lindman and Savage [1963] and Smith [1969]. Nevertheless, Bayesian methods seem to be slow in catching on in psychology and other fields. Whether this is due to institutional inertia, lack of textbooks, or to more fundamental difficulties in matching a Bayesian approach to research contexts is at this point a matter of judgment. It is true that good examples of attempts systematically to apply Bayesian methods to actual research problems in psychology and the social sciences are almost non-existent.

8 *Conclusion.*

As stated earlier, the value of *The Significance Test Controversy* for philosophers is that it provides a good source of problems that any adequate account of inductive reasoning and scientific method should be able to resolve. Let anyone who thinks he has such an account try his hand here. But one cannot stop here. To develop an adequate comprehensive account of scientific inference one will eventually have to examine some actual cases of psychological and sociological research. There might even be more to learn from such 'case studies' than from the explicitly methodological reflections of the scientists involved.

A final comment on the text. No one should attempt to read this volume straight through. It is sufficient to begin with the most important, comprehensive or clearest selections. The papers by Selvin, McGinnis, Kish, Morrison, Winch, Rozeboom and Bakan are particularly recommended. Those with some background in the foundations of statistics may want to dig for further variations on dominant themes. Try the papers by Kendall, Davis, Gold (1958), Bishers, Camilleri, Skipper, Meehl and Lykken. Those without much background in statistical theory should go back to the first selection from Hogben's [1957]. In fact, Hogben's book is an excellent introduction to the whole subject area. After that one would be prepared to tackle the primary sources, *e.g.* Fisher, Neyman-Pearson, Neyman, Wald, Savage, and also secondary sources, *e.g.* Barnard, Cox, Tukey, Birnbaum, Hacking.[1] One may hope that in time philosophers will appreciate the significance of the significance test controversy for inductive logic and that it will begin to appear in the philosophical literature. Eventually we may have something relevant to say to sociologists and psychologists who, it seems, very much want and need to hear something relevant from someone who understands the full dimensions of the problem.

RONALD N. GIERE
Indiana University

---

[1] In addition to references in previous footnotes, cf. Fisher [1956], Neyman [1950], Savage [1954], Barnard [1949], Cox [1958], Tukey [1960] and Birnbaum [1969].

REFERENCES

ARBUTHNOT, J. [1710]: 'An Argument for Divine Providence Taken from the Constant Regularity Observed in the Birth of Both Sexes', *Philosophical Transactions of the Royal Society of London*, **27**, pp. 186–90.

BARNARD, G. A. [1949]: 'Statistical Inference', *Journal of the Royal Statistical Society* (*B*), **11**, pp. 116–49.

BIRNBAUM, A. [1967]: 'John Arbuthnot (1667–1735), Scholar-Scientist, Good-Humoured Satirist, and Deviser of the Earliest Statistical Test of Significance', *The American Statistician*, **21**, Number 3, pp. 24–5.

BIRNBAUM, A. [1969]: 'Concepts of Statistical Evidence', in S. Morgenbesser, P. Suppes and M. White (*eds.*): *Philosophy, Science and Method* pp. 122–43.

COX, D. R. [1958]: 'Some Problems Connected with Statistical Inference', *Annals of Mathematical Statistics*, **29**, pp. 357–72.

CRAMÉR, H. [1946]: *Mathematical Methods of Statistics.*

EDWARDS, W., LINDMAN, H. and SAVAGE, L. J. [1963]: 'Bayesian Statistical Inference for Psychological Research', *Psychological Review*, **70**, pp. 193–242.

FISHER, R. A. [1925]: *Statistical Methods for Research Workers.*

FISHER, R. A. [1935]: *The Design of Experiments.*

FISHER, R. A. [1956]: *Statistical Methods and Scientific Inference.*

GIERE, R. [1971]: 'Objective Single Case Probabilities and the Foundations of Statistics', forthcoming in P. Suppes (*ed.*): *Proceedings of the Fourth International Congress for Logic, Methodology and the Philosophy of Science.*
HACKING, I. [1965]: *Logic of Statistical Inference.*
HAYS, W. H. [1963]: *Statistics for Psychologists.*
HEMPEL, C. G. [1962]: 'Deductive-Nomological vs. Statistical Explanation', in G. Maxwell and H. Feigl (*eds.*): *Minnesota Studies in the Philosophy of Science*, **3**, pp. 98–169.
HODGES, J. L. and LEHMAN, E. L. [1954]: 'Testing the Approximate Validity of Statistical Hypotheses', *Journal of the Royal Statistical Society* (*B*), pp. 261–8.
HODGES, J. L. and LEHMAN, E. L. [1964]: *Basic Concepts of Probability and Statistics.*
HOGBEN, L. T. [1957]: *Statistical Theory: The Relationship of Probability, Credibility, and Error. An Examination of the Contemporary Crises in Statistical Theory from a Behaviourist Viewpoint.*
LAPLACE, S. [1814]: *Philosophical Essay on Probabilities.*
LEHMAN, E. L. [1959]: *Testing Statistical Hypotheses.*
LEVI, I. [1967]: *Gambling with Truth.*
NEYMAN, J. [1950]: *First Course in Probability and Statistics.*
NEYMAN, J. [1952]: *Lectures and Conferences on Mathematical Statistics and Probability.*
NEYMAN, J. and PEARSON, E. S. [1967]: *Joint Statistical Papers.*
POPPER, K. R. [1967]: 'Quantum Mechanics without the Observer', in M. Bunge (*ed.*), *Quantum Theory and Reality*, pp. 7–44.
SAVAGE, L. J. [1954]: *The Foundations of Statistics.*
SMITH, C. A. B. [1969]: *Biomathematics.*
TUKEY, J. [1960]: 'Conclusions versus Decisions', *Technometrics*, **2**, pp. 423–433.

BAYESIANISM AND THE RATIONALITY OF SCIENTIFIC INFERENCE[1]

Wesley Salmon's concluding paragraphs[2] begin

> The analysis of the inference by which scientific hypotheses are confirmed by observational evidence shows, I believe, that its structure is given by Bayes's theorem. This schema provides a place for the hypothetico-deductive method, showing that it is fallacious in its crude form, but that it can be made into a valid method when appropriately supplemented. Two kinds of probabilities are needed to supplement the hypothetico-deductive schema. We must assess the probability that our observational results would obtain even if the hypothesis under consideration were false. For strongest confirmation, this probability should be small. This seems a natural interpretation of Popper's methodological requirement that scientific hypotheses must be audacious and take risks. We must, in addition, assess the prior probabilities of the hypotheses we are considering. This is a reasonable interpretation of Hanson's demand for plausibility arguments.

With this I am in full agreement. But my Bayesianism is personalist[3] whereas

---

[1] Review of *Essays in Contemporary Science and Philosophy* (*University of Pittsburgh Series in the Philosophy of Science*) *ed.* R. G. Colodny: volume 3, *Mind and Cosmos* (Pittsburgh, 1966), 362 pp., £3.20; volume 4, *The Nature and Function of Scientific Theories* (Pittsburgh, 1970), 361 pp., £6.14. I shall devote extended attention to parts of Salmon: 'The Foundations of Scientific Inference', pp. 135–275 in volume 3, and to Shimony: 'Scientific Inference', pp. 79–172 in volume 4; the remaining contents of the volumes will be more briefly surveyed.

[2] Other aspects of Salmon's monograph (*op. cit.*, *above*, footnote 1) have been reviewed by I. Levi (Levi [1968*a*]). It has been issued separately as a paper-back.

[3] For a general account of the theory of personalism and how it has solved the problem of induction, cf. Hacking [1966] and Gillies [1972]. 'Personalist' and 'Subjectivist' are used synonymously in the literature.

Salmon's remains frequentist. His objections to full-blooded personalism merit scrutiny.

Salmon objects that the personalist requirements on scientific rationality are too liberal. But one might equally object that mathematical logic is too liberal. It too allows one to start with any consistent set of premises provided one's subsequent inferences are valid. The objection would be serious only if the personalist or the logician insisted on rejecting additional assumptions of the same degree of evidence as his own. But in fact no such additional assumptions have been found and for a sufficiently important class of applications it can be shown that no such additional assumptions are needed.

However Salmon does also raise the original and specific objection that personalism may let in counter-inductive strategies[1]; though he does not give an example of an actual prior probability distribution over the universe which would legitimate a counter-inductive strategy. Such examples as I have been able to construct seem to be coherent but not strictly coherent.[2] However, even if a genuine example could be constructed, I would be quite happy to know that only a vanishingly small proportion of all prior probability distributions over the universe legitimated counter-inductive strategies rather than inductive ones. If Salmon is still hankering for a stronger justification of induction than this, then I think he is hankering for too much.

Salmon thinks that the most fundamental objection to the personalist model of scientific inference is that it allows too much arbitrariness in the choice of the priors that are assigned to hypotheses and that it fails to assign an over-riding role to experience in the determination of the choice of these priors. But to my mind the beauty of the personalist approach is that it does always allow us to push the argument successive stages further back, to yet earlier priors, thereby enabling us to take into account any amount of evidence relevant to the determination of the original priors. It seems to me that this kind of thorough-going personalism (especially cogently presented in Levi's important review[3] of Hintikka and Suppes's book) does automatically take into account all the considerations which Salmon thinks should guide us in the choice of priors. Pushing the argument successive stages further back will determine the original priors more and more precisely (by reducing disagreement over them) provided merely that the successive priors assigned by rival theorists *do not diverge ever more widely* as we go further back. But in the scientific inference case, the condition of the proviso is assured so long as the beliefs of the scientists may ultimately be traced back to the beliefs of common sense. Of course if Salmon hopes ultimately to eliminate all arbitrariness in the choice of priors, and hence to eliminate scientific disagreements altogether, then it seems to me that he is chasing a will-o'-the-wisp.

---

[1] I take it that Salmon means general counter-inductive strategies. For it is easy to think of examples of particular properties for which we must all admit that the probability of n individuals possessing that property is a rapidly decreasing function of n, and here we are obviously all committed to counter-inductive strategies.

[2] Thus if one assumed that for any sufficiently simple property A, exactly half the objects in the universe had the property A and half the property $-A$, then to legitimate counter-induction one would have to assume *either* a known finite size for the universe, *or* that the probability of there being n objects in the universe were a rapidly decreasing function of n. But while both these assumptions are coherent, neither is strictly coherent.

[3] Levi [1968*b*].

It is Abner Shimony who, in the subsequent Pittsburg volume, takes up the banner of the Bayesian approach to scientific inference. But, although himself no frequentist, Shimony shares Salmon's reluctance to accept the full personalist position. I shall examine the sources of Shimony's reluctance and the doubts which lead him to propose a monster which he calls 'tempered personalism'.

First there is the familiar objection that the rational degree of credibility associated with the literal and unqualified truth of a physical theory, in conditions far different from those in which experiments have been conducted, *e.g.* in arbitrarily small space-time regions, throughout all space and time, and so on, must be zero or extremely small. Shimony suggests in answer to this objection that we should associate a non-negligible probability, not with the theory h itself, but with the conjunction of the propositions:

(i) within the domain of current experimentation h yields almost the same observational predictions as the true theory;
(ii) the concepts of the true theory are generalisations or more complete realisations of those of h;
(iii) among the currently formulated theories competing with h, there is none which better satisfies conditions (i) and (ii).

Shimony's answer here seems to me unnecessarily elaborate. In analysing any particular inference according to the Bayesian schema we will necessarily have constructed a mutually exclusive and jointly exhaustive set of hypotheses, so (iii) is redundant. (ii) seems to me unworkable. Does Shimony consider that the concepts of quantum mechanics, or of special relativity, or of quantum electrodynamics, or of general relativity, are 'generalisations or more complete realisations' of the concepts of classical mechanics, of Newtonian kinematics, of Maxwellian electromagnetic theory, or of Newtonian gravitational theory? He gives us no clue as to how to answer such questions.

Moreover I am not convinced that we need explicitly impose any of Shimony's conditions provided we ascribe probabilities directly to scientific hypotheses as understood by scientists. What is important is not how philosophers construe physicists' theories but how physicists construe them. But it is philosophers, not physicists, who take a theory to be asserted as true for arbitrarily small space-time regions, throughout all space and time, for arbitrary extremes of conditions whose variation has not been studied experimentally, and so on, unless anything is explicitly said to the contrary. Physicists would normally assume that no such commitment is included unless it be explicitly asserted. So is there really a difficulty of the kind envisaged, when one comes to consider actual examples of theoretical issues on which the progress of physics has depended and of the sort which it would be instructive to subject to Bayesian analyses? Whether or not Venus's orbit encircles the sun? Whether or not the earth has an annual motion relative to the fixed stars? Whether or not the smaller parts of a gas apparently at rest are in reality in violent motion? Whether contiguous portions of palpably homogeneous bodies cease to be similar when their dimensions are of the order of a thousand-millionth of a centimetre? Whether or not the time-interval between two events is independent of the path along which it is measured? Whether a man falling in a lift could readily detect that his motion was non-inertial by observing the trajectory of a light ray? Whether energy is conserved in the processes of emission and absorption of light? Whether or not energy is conserved in beta-decay? Whether the four-fermion interaction contains equal amounts of vector and axial vector coupling? There seems no *prima facie* reason why the rational degree of belief in such propositions or their negations should always be zero or extremely small. Of course if, in deference to L.S.E. philosophy, we graft into such propositions implicit universal quantifications with respect to additional unmentioned variables, then we can no doubt persuade ourselves that now the probabilities must be zero or extremely small. But surely the study of scientific inference is concerned with the theories that scientists discuss and especially with the parts that scientists sometimes come to believe in, not with the theories invented by philosophers that no one could ever possibly believe in.[1]

A more serious difficulty, according to Shimony, for the Bayesian, is that the personal probabilities associated with scientific hypotheses cannot be explicated in terms of betting ratios, since bets on scientific hypotheses can never be conclusively settled. I agree and would add 'either way' after 'settled'.[2] The difficulty creates two problems: first that of providing an alternative explication for personal probabilities to that in terms of betting ratios,[3] second that of providing a justification for the requirement that such probabilities should satisfy the probability axioms. Obviously the second problem cannot be solved without solving the first problem. Nevertheless Shimony attempts to do so. He tries to do this by making as weak assumptions as possible concerning the measures of 'rational degrees of commitment' (his name for the personal probabilities associated with scientific hypotheses), but ones which are nevertheless strong enough to enable him to derive the probability axioms. In general his derivation[4] follows earlier ones of Aczél, Cox, and Good. But since, the way Shimony presents it, the concept of 'rational degree of commitment' is merely his theoretical device for explicating scientific inference, and is not required to play any other role, he would apparently be entitled to assume any axioms he liked concerning it. What is chiefly missing here seems to be a requirement concerning the role that our totality of personal probability ascriptions to scientific hypotheses has to play in determining our personal probability ascriptions to many ordinary decidable propositions: in other words a requirement concerning the point of the scientific enterprise. The fact that the latter ascriptions must (for the usual Dutch book reasons) satisfy the probability axioms presumably constrains the former to do so also.

---

[1] My view involves admitting that the degree of generality of the former is usually rather vague and ill-defined though roughly ascertainable from the context. (In science, unlike philosophy, a certain amount of vagueness and lack of definition is an advantage in the right places.) The consequential numerical imprecision in the subjective probabilities is by no means fatal to Bayesianism.

[2] Shimony seems reluctant to add this. But it seems to me that unless this is insisted upon, one can easily define the probability of a scientific hypothesis, in terms of betting ratios, as the limit to which the betting ratio on a negative outcome of n potential falsifying experiments tends, as n tends to infinity, in an enumeration of all potential falsifying experiments for the hypothesis in question. This is the standard and obvious way of interpreting the personal probability associated with a generalisation concerning infinitely many outcomes each of which is individually decidable. I take it that we do not adopt it here only because we also believe that if e is observational evidence and h a scientific hypothesis, $P(e, h)$ is never strictly zero, or in other words that however improbable the observed facts may seem, given a particular scientific hypothesis, they can never logically oblige a sufficiently imaginative and stubborn defender to abandon it.

[3] Or that in terms of utility theory. For it is just as difficult to associate a utility with a non-outcome as it is to bet on a non-outcome. If the difficulty *merely* concerned direct explication of personal probabilities in terms of bets, then the personalist could legitimately reply that personal probabilities acquire meaning through their role in utility theory and that he has no more obligation than any other scientist to provide explicit operational definitions for his theoretical entities.

[4] I have two small reservations about Shimony's derivation. First it seems to me that he needs to prove the uniqueness and not merely the existence of the probability function he constructs. Second he seems to make unnecessarily heavy weather over the continuity of functions which only need, it seems, to be assumed continuous where, if anywhere, the domain of definition of their original arguments is continuous.

Nothing I have said of Shimony's position, so far, really puts him outside the camp of orthodox personalism. However the way he deals with a third objection to personalism does so. I shall try to understand this objection in a moment; but first here is his proposed modification of personalism. He proposes that it is only with some definite set of envisaged observations in mind that we should ever form a set of prior probabilities, and that these should satisfy the condition that

> the prior probability of each seriously proposed hypothesis must be sufficiently high to allow the possibility that it will be preferred to all rival, seriously proposed hypotheses as a result of the envisaged observations.

Shimony calls the result of applying this restriction 'tempered personalism'. Shimony emphasises the point that decisions about classifying hypotheses as seriously proposed or not are not to be 'left entirely to the subjective judgement of individual investigators' and here he urges, *inter alia*, respect for the scientific establishment.

Let us consider the application of Shimony's proposal to some examples. Suppose an experimental programme in high energy physics led to the detection of massive particles of charges 1/3 and 2/3. The personalist would infer a high posterior probability for something like the 3-quark model of Gell-Mann and a low posterior probability for rival quark models, as well as for all no-quark models. Not so, apparently, the tempered personalist: he will presumably infer nothing at all so long as the experimental programme in question would have been incapable of yielding a result which led us to prefer (by yielding a posterior probability of greater than one half for) any particular one of the rivals to the 3-quark model. Similarly if someone finds, in the sludge at the bottom of the Pacific, particles which behave very like magnetic monopoles, personalists may conclude that magnetic monopoles probably exist; tempered personalists will apparently infer nothing unless a negative result would have satisfied them that it was more probable than not that the universe contains no magnetic monopoles. Something seems to have gone wrong here.

Would Shimony's proposals be satisfactory if he weakened his tempering condition to the simple requirement that all seriously proposed hypotheses be assigned non-negligible priors? I think not. At a rough estimate at least fifty per cent of all hypotheses seriously proposed in current physics journals prove on closer inspection to be internally incoherent, so why should I assign them non-negligible priors? Suppose the majority of the remainder rest on blatantly fallacious arguments by their proponents which merely *purport* to show that they have non-negligible priors. The fact that a Wheeler, or a Heisenberg, or a Bondi, or a Dicke, has persuaded himself on the basis of some appalling *non-sequitur* that a certain hypothesis has a substantial prior probability, seems to me no reason why I should not, on the basis of an honest independent assessment of the strength of the scientist's arguments, assign his hypothesis a negligible prior probability. I would have thought that scientists had to make up their own minds about other scientists' arguments; indeed that this is partly what science is all about.

How then does Shimony come to feel a need for his tempering condition? It is apparently that he wishes to rule out as irrational what he calls 'prejudice'. Now if he is to be successful in this he must evidently first succeed in defining 'prejudice' in such a way that it is not already ruled out by ordinary personalism and in such a way as to embrace only things we all regard with disfavour. But Shimony does not attempt to define 'prejudice'. At one point he seems to imply that any opinions are prejudiced which deviate sufficiently from those of the generality of mankind seriously to delay the attainment of general consensus:

> If personal probabilities were used, unqualified by the tempering condition, then deep-seated prejudice . . . could produce immense (and possibly infinite) prior probability ratios . . . and consensus would be severely delayed (p. 103)

But why does he regard this consequence as objectionable? (Perhaps there is a hint in his admission in a footnote that he would have liked to have called tempered personalism 'socialism'.)

Elsewhere (p. 92) Shimony implies that a man should be deemed prejudiced against a hypothesis if he does not accept it when there is 'a very large body of favourable observational data'. This criterion would presumably condemn as prejudiced those who do not accept ghosts, ESP, flying saucers, or the Copenhagen interpretation of quantum mechanics.

Thus neither Shimony's tempering condition, nor what he actually says about prejudice, give us any very clear idea of what it is that he really wishes to rule out. In particular he does not offer us any concrete examples of kinds of prejudice that would be let in by orthodox personalism and which he would wish to eliminate. Kyburg has given[1] a more precise description of kinds of prejudiced behaviour which he thinks strict personalism still allows, but Shimony does not seem to have Kyburg's sort of example in mind, and anyway Kyburg's examples can in fact be ruled out by Dutch book considerations involving conditional bets.[2]

I think Shimony's deeper reason for attempting to temper personalism must be that he shares Salmon's, already criticised, belief that personalism imposes no practical constraints on the scientist's choice of priors. But this is hardly so. And if one considers particular examples I think one will find that the man we all agree is prejudiced is perfectly prepared to assign reasons for his choice of priors; that these reasons can generally be couched in the form of Bayesian inferences from earlier priors (deductive inferences are, after all, just a special case of Bayesian inferences); and that these earlier priors are unimpeachable; but that either the evidence cited is imaginary or there are identifiable fallacies in the inferences by which the later priors have been derived from the earlier ones. At any rate before orthodox personalism is condemned here, I think we are entitled to some actual examples of the kinds of scientific prejudice which it is deemed by Salmon and Shimony to be itself incapable of condemning.

---

[1] Cf. his [1970], pp. 73–74.

[2] If we know a man is prejudiced in the sense that he is going, if and when the evidence comes in, systematically to depart from Bayesian conditionalisation in a certain direction, then we can certainly bet with him so as to have no chance of losing and a finite chance of winning money off him. Dutch book arguments are thus not restricted merely to establishing coherence at each particular time *t*.

There is a further consideration which is perhaps not sufficiently borne in mind by those who wish to impose severer constraints on the choice of priors than those implicit in personalism itself. This is that given any hypothesis, however incredible it may be, its subjective probability for any individual is calculable provided that individual can envisage a sequence of possible observations, which, if it occurred, would eventually convince him that the hypothesis in question had become more probable than not. If we assume for simplicity that the probability of the observations in question given the hypothesis is very close to unity (of course for no scientific hypothesis will the probability of any observations be exactly unity), then the probability the subject attributes to the observations in question (which must be finite, because of strict coherence) suffices to determine the prior probability he must attribute to the hypothesis. It is therefore not the case that any hypotheses which may come to be accepted later can at a given time have probabilities too small to estimate, contrary to the impression given by some of personalism's critics. This seems one more reason why the personalist need not treat seriously proposed hypotheses with any special favour or disfavour.

I can only conclude this discussion by remarking that it seems a pity that the beauties of a strictly Bayesian approach to the problem of the rationality of scientific inference should have been presented to the philosophical public in so half-hearted a form by such distinguished authors.

Salmon and Shimony's articles are each the most substantial, as well as, for me, the most interesting, in their respective volumes. I was rather disappointed with the other philosophers in volume 3. Carl Hempel's 'Recent Problems of Induction' seemed all very familiar stuff. Dudley Shapere, on 'Meaning and Scientific Change' is most remarkable for his negative recommendations. He encourages us to leave our analytical tools behind, abandoning such distinctions as those between 'meaningful' and 'meaningless', 'scientific' and 'unscientific', 'empirical' and 'non-empirical', and concentrate on the detailed study of case-histories. Sylvain Bromberger, on the other hand, feels no temptation, in discussing 'why-questions', to sacrifice analytical rigour to scientific relevance. His paradigm of a 'why-question' and of a satisfactory answer to such a question is ' "Why is the plural of the French noun *cheval*, *chevaux*?" . . . Answer: "Because *cheval* ends in *al*." ' His analysis is nevertheless important, though not in proportion to the number of grants he was awarded for the work.

Another attempt to replace Hempel and Oppenheim's necessary conditions for an explanation by ones that are also sufficient is Salmon's 'Statistical Explanation' (which is not exclusively concerned with the statistical kind) in volume 4. I was more sympathetically disposed to this, since Salmon's approach is essentially Bayesian. However it did seem that his claim that while it is a sufficient condition for an explanation, that it should raise the antecedent probability of the explanandum, it is not necessary that it should render the explanandum actually probable, is only true in some contexts. Thus what demands we make of an explanation seems to be discipline-dependent. If this is so philosophers could either opt for loose Hempelian criteria of explanation, or they could single out as paradigmatic some rather strict criterion of explanation and then consider to what extent this criterion is, is not, and could in principle be, satisfied in different disciplines. Salmon and Bromberger seem both to fall between these two stools, proposing criteria which seem too weak in some contexts, too strong in others. And both authors fight shy of the deeper examples of scientific explanations.

Norwood Russell Hanson's 'A Picture Theory of Theory Meaning', in volume 4, is almost on the same subject being on what it is to understand a scientific theory and what it is to understand nature by means of such a theory. Hanson's detailed illustrations from aerodynamic theory are beautifully presented, even if the questions he is interested in seem sometimes to belong properly to the psychology of science rather than to the philosophy of science.

Herbert Simon is a professional in the psychology and artificial intelligence of problem solving. His articles 'Thinking by Computers' and 'Scientific Discovery and the Psychology of Problem Solving', both in volume 3, are philosophically sensitive as well as being informative. He offers us some useful expert tips on how to solve general problems. If anything he seems to overplay the rational aspects of scientific discovery. It is surely a historical fact that sheer muddle has played a not inconsiderable rolė in the development of physics. This is, to my mind, one of the most striking features of the writings of a Maxwell, a Planck, a Bohr, or even, on occasions, an Einstein. But perhaps this is a historical accident rather than of the nature of the process.

In 'Cosmic Processes and the Nature of Time', another scientist contributor to volume 3, Thomas Gold, gives one of his splendid expositions of the problems concerning Time's arrow. This time Gold's main aim is to offer an explanation of our sense of the present moment travelling through space-time. He succeeds in convincing the reader that the existing asymmetry of time is crucial to our possession of this conviction, though he omits to notice that his arguments really only establish that the asymmetry is a necessary condition, not that it is also a sufficient condition, for our sense of the present moment travelling through time.

Of the three remaining articles in volume 3, Henry Margenau's 'The Philosophical Legacy of the Quantum Theory' is not among his better contributions to the subject, and Joseph Clark's 'The Physiognomy of Physics' and 'Science and Some Other Components of Intellectual Culture' seem chiefly remarkable for the amount of alliteration in which he indulges.

In volume 4, Grover Maxwell's 'Theories, Perception, and Structural Realism' is partly a sort of Bayesian approach to epistemology. He writes:

> I contend that realism is much more highly confirmed than any other theory that 'explains the facts', because its *prior probability* is so much greater than any other such theory, and prior probability is always one of the crucial factors in selecting among competing hypotheses all of which explain current evidence. My reasons for accepting realism are of the same kind as those for accepting any scientific theory over others which also explain current evidence.

However I am uneasy about this application of Bayesianism. I would need to be convinced first that realism is indeed an *empirical* theory (by having conceivable observations described to me which would, if they occurred, persuade me that realism was after all probably false, rather than that I was dreaming those observations, or that the external world had merely suddenly ceased to exist having existed up to that moment). Second it seems to me that the fact, that in the epistemological, unlike the scientific case, the prior probabilities assigned would have to be ultimate, and not further analysable, makes a Bayesian justification here suspiciously question-begging.

Mary Hesse's 'Is There an Independent Observation Language' seemed to me to rely on dubious arguments where it was attacking positions which are still of interest. Her argument that there are no observational predicates since even applications of the predicate 'red' must presuppose laws (and hence it must count as a theoretical predicate), since such applications are publicly correctible, seemed to me too brief. Why should not presupposed conventions, rather than presupposed laws, be sufficient for public correctibility? Again in attempting to show that there can be no predicates which are purely theoretical in the sense of only being applicable to theoretical entities,[1] she seems to presuppose that a new predicate can only be introduced into a language either ostensively, or through explicit definition as a combination of other predicates. Least satisfactory of all seemed her attempt to show that so-called correspondence rules, of which her example is ' "This exerts strong repulsive forces" implies "This is hard",' are synthetic rather than analytic statements. To show this she would evidently have to show that the proposition in question is not among the consequence class of the statements formed by the union of the meaning postulates of the Newtonian theory of forces and the analytic statements of ordinary language. But, amazingly, Hesse considers it sufficient to observe that ' "repulsion" and "hardness" are not synonymous in ordinary language." ' I thought her other papers in this area better than this one.

Finally, in 'Problems of Empiricism, Part II', Paul Feyerabend offers us an entertaining irrational reconstruction of Galileo, which will surely become a classic in the secondary literature on seventeenth-century science. It is not that Feyerabend tries to give rational reconstructions of Galileo's arguments but fails; Feyerabend does not of course even try. For him the purpose of scientific theory is aesthetic and it is rhetorical arguments, not logical ones, that are called for when a novel theory is to be defended. Nevertheless there are others of us for whom the purpose of a scientific theory is primarily to make predictions which, given everything else we know and suspect, seem likely to prove correct. We are entitled to assess the arguments by which a scientific theory is defended from this point of view too. The Bayesian approach to scientific inference provides a precise formal framework for doing this and one which is perfectly consistent with many of Feyerabend's points (especially with those where he agrees with Lakatos) in particular with the non-existence of logically crucial experiments, with the facts that a theory can live with many *prima facie* refuting instances, and that the auxiliary theories needed to derive a theory's claimed observational successes may themselves for long be merely pious hopes, and so on. Of course the Bayesian approach could hardly be reconciled with all of Feyerabend's theses. But he himself seems already elsewhere to have retreated considerably from the earlier extremer versions of the incommensurability thesis, and his exaggerated empirical psychological thesis concerning the extent of the determination of people's perceptions by their beliefs and desires (which he toys with in the usual extreme form in the present paper), had I thought been reduced to its proper proportions in the proper place, namely the experimental psychological literature, years before Hanson, Kuhn, and Feyerabend chose to revive it in the philosophical literature. I would not deny for one moment that, given the evident weakness of the logical faculties of most men, and *inter alia* of many men of science, much purely rhetorical and not rationally reconstructible argument has played an important part in the history of science. But one would have thought that the philosopher of science would be most interested in picking out and analysing in detail those scientific arguments which did seem to be rationally reconstructible.

---

[1] Predicates which immediately came to my mind were 'is an exact differential', 'is antisymmetric', 'is unitary', 'is divergence-free'.

JON DORLING
Chelsea College of Science and Technology

REFERENCES

Gillies, D. [1972]: 'The Subjective Theory of Probability', *British Journal for the Philosophy of Science*, **23**, pp. 138–157.

Hacking, I. [1966]: 'Subjective Probability', *British Journal for the Philosophy of Science*, **16**, pp. 334–9.

Kyburg, H. E. [1970]: *Probability and Inductive Logic.*

Levi, I. [1968*a*]: Review of W. C. Salmon: 'The Foundations of Scientific Inference', *British Journal for the Philosophy of Science*, **19**, 259–61.

Levi, I. [1968*b*]: Review of J. Hintikka and P. Suppes (*eds.*): 'Aspects of Inductive Logic', *British Journal for the Philosophy of Science*, **19**, pp. 73–81.

Law and Explanation

An Essay in the Philosophy of Science

Peter Achinstein

The aim of this book is to examine the concept of scientific law, one of the most important in the sciences. The author considers those characteristics of laws in virtue of which they express regularities, features that constitute their generality and necessity, the nature of explanation and the kinds of explanations laws supply, and types of reasoning in which the scientist engages when he proposes a law.
£2.75

The Philosophies of Science

H. R. Harré

This book shows how various views about the nature of science are related to the great historical schools of philosophy. The argument is set out in terms of certain concrete episodes in the history of science—a manner of exposition which brings out most clearly the influence of philosophical theories on the development of science, and of scientific discovery on modes of thinking in philosophy.
7 text figures £1.30 paper covers 60p OPUS

A Historical Introduction to the Philosophy of Science

John Losee

Since the time of Plato and Aristotle philosophers have raised questions about the proper evaluation of scientific interpretations. This book is a historical exposition of positions that have been held on such issues as the distinction of observation and theory, and the evaluation of competing scientific theories.
26 text figures £2 paper covers 90p OPUS

Minnesota Studies in the Philosophy of Science Volume V

Historical and Philosophical Perspectives of Science

Edited by **Roger H. Stuewer**

The essays which make up this volume are for the most part based on papers presented or discussed at a conference which explored the relationships between the history and the philosophy of science. £5.50 *University of Minnesota Press*

OXFORD UNIVERSITY PRESS

Mathematical Logic

with special reference to the natural numbers

S. W. P. STEEN

This book presents a comprehensive treatment of basic mathematical logic. The author makes exact the vague, intuitive notions of *natural number*, *preciseness*, and *correctness*, and invents a method whereby these notions can be communicated to others and stored in the memory.

£15.00 net

The Evolution of Physics

ALBERT EINSTEIN and
LEOPOLD INFELD

The layman does not often have the opportunity of reading a simple exposition of advanced scientific thought written by the men who did the actual creative thinking. In this book there will be found an easily understood but authoritative account of the growth of ideas in physical science from the earliest concepts to the more abstruse theories of modern times.

Hard covers £3.60 net
Paperback £1.60 net

Likelihood

an account of the statistical concept of likelihood and its application to scientific inference

A. W. F. EDWARDS

An account of how likelihood can be used to provide an acceptable measure of support for scientific hypotheses in the light of observational and experimental data which bear upon them. Estimation, tests of significance, randomization, experimental design, and many other statistical topics are explored using the likelihood approach. £3.80 net

CAMBRIDGE
UNIVERSITY PRESS

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY

Edited by Marvin Farber

MARCH, 1972, VOLUME XXXII, No. 3

CONTENTS

I ARTICLES
- The Logic of Ethical Statements, by ROY STONE de MONTPENSIER
- Life-World Constitution of Propositional Logic and Elementary Predicate Logic, by L. ELEY
- Other Minds and Professor Ayer's Concept of a Person, by P. M. S. HACKER
- Karl Marx and the Rights of Man, by BETTY A. SICHEL
- Max Scheler's Phenomenology of Shame, by PARVIS EMAD

II DISCUSSIONS AND COMMUNICATIONS

III REVIEWS

IV NOTES AND NEWS, RECENT PUBLICATIONS

Subscription prices $7.50 per year for individuals.

$9.00 per year for institutions.

Separate issues, $2.00 and $2.50 respectively.

Postage and handling for back issues only.

Papers submitted for publication will not be returned unless accompanied by a self-addressed stamped envelope, or equivalent postage.

A circular listing the main contents of the Journal since 1940 will be sent upon request.

Address all communications to
Philosophy and
Phenomenological Research
State University of New York at Buffalo
Buffalo, New York 14226

Studies in the Foundations, Methodology
and Philosophy of Science

Volume 4

Problems in the Foundations of Physics

Edited by Mario Bunge

With 3 figures
VII, 162 pages 1971
Cloth DM 54,–
ca. US $ 15.60

This book consists of 11 original papers dealing with problems in the foundations and the philosophy of physical science. "Physical science" is conceived in the large sense, as embracing all the sciences concerned with the inorganic world. The problems tackled in this volume concern a number of theories, classical, relativistic, and quantal, studied everywhere but seldom subjected to a deep scrutiny. These problems are of permanent interest, for they concern such basic concepts as position and probability, or long term research problems like quantization and the casting of the basic laws in the form of variational principles.

Contents:

Part I: General: M. Bunge: A Mathematical Theory of the Dimensions and Units of Physical Quantities. – D. G. B. Edelen: Reflections on Variational Principles and Invariance Theory. – **Part II: Of Particles:** P. Havas: Galilei- and Lorentz-Invariant Particle Systems and Their Conservation Laws. – **Part III: Of Fields:** P. G. Bergmann: Foundations Problems in General Relativity. – E. J. Post: Geometry and Physics: A Global Approach. – **Part IV: Of Quanta:** A. Fine: Probability in Quantum Mechanics and in other Statistical Theories. – A. J. Kálnay: The Localization Problem. – A. Komar: Semantic Foundations of the Quantization Program. – **Part V: Of Molecules:** R. Aris: Algebraic Aspects of Formal Chemical Kinetics. – **Part VI: Philosophical:** S. A. Kleiner: Criteria for Meaning Changes in Theoretical Physics. – T. W. Settle: The Relevance of Philosophy to Physics.

Published earlier:

Volume 1: Delaware Seminar in the Foundations of Physics
Edited by Mario Bunge. With 5 figures. XII, 193 pages. 1967. Cloth DM 42,–; ca. US $ 12.20

Volume 2: Quantum Theory and Reality
Edited by Mario Bunge. VI, 117 pages. 1967. Cloth DM 33,–; ca. US $ 9.60

Volume 3: Mario Bunge, Scientific Research. Two Parts (sold separately)
Part I: The Search for System
With 64 figures. XII, 536 pages. 1967. Cloth DM 92,–; ca. US $ 26.60
Part II: The Search for Truth
With 61 figures. VIII, 374 pages. 1967. Cloth DM 75,–; ca. US $ 21.70

Springer-Verlag
Berlin · Heidelberg · New York
München · London · Paris · Tokyo · Sydney

Interpretations of Life and Mind

Essays around the Problem of Reduction

EDITED BY MARJORIE GRENE
Professor of Philosophy,
University of California, Davis

These papers represent a combined attack by both philosophers and scientists on the tendency of modern science to reduce biology to physics, the person to a nexus of behaviour or minds to machines. They examine critically the principles and implications of reductionism and work towards a viable alternative. The contributors are Ilya Prigogine, Marjorie Grene, Charles Taylor, Anthony J. P. Kenny, Amélie Rorty, Alasdair C. MacIntyre, Hubert Dreyfus, William T. Scott, Robert Cohen. £2.25

ROUTLEDGE
London and Boston

British Journal for the Philosophy of Science

is one of 49 learned journals published by Cambridge University Press.

A descriptive catalogue of Cambridge journals including details of back volumes is available from the London and New York offices of the Press.

Cambridge journals may be ordered from a bookseller or direct from the publishers.

Cambridge University Press
Bentley House, 200 Euston Road
London,NW1 2DB

American Branch: 32 East 57th Street
New York, N.Y. 10022

FOUNDATIONS OF LANGUAGE

International Journal of Language and Philosophy

Editorial Board: **Morris Halle (M.I.T.), Peter Hartmann (Konstanz), K. Kunjunni Raja (Madras), Benson Mates (Univ. of Calif.), J. F. Staal (Univ. of Calif.), Pieter A. Verburg (Groningen), John W. M. Verhaar (Djakarta).**

CONTENTS OF VOLUME 7, NO. 4, NOVEMBER 1971

Derek Bickerton / Inherent Variability and Variable Rules—John B. Walmsley / The English Comitative Case and the Concept of Deep Structure—Gift Siromoney / Grammars for Kernel Sentences in Tamil—Barry Richards / Searle on Meaning and Speech Acts.

Discussion: Semantics and Semantics (Bruce Vermazen).

Review Article: Stress in Arabic and Generative Phonology (Michael K. Brame).

Reviews: Rosane Rocher / *La théory des voix du verbe dans l'école pāṇinéenne* (Le 14e āhnika) (B. A. van Nooten)—E. D. Hirsch, Jr. / *Validity in Interpretation* (J. J. A. Mooy) —Index of names.

Subscription price per volume of four issues Dfl. 115 (US $37.38), including postage.
One volume is published yearly
Personal subscription price on request

D. REIDEL PUBLISHING COMPANY / DORDRECHT - HOLLAND

Revue Internationale de Philosophie

Direction Scientifique: Professeurs Ph. DEVAUX, H GOUHIER, et Ch. PERELMAN

Administration: Mme Simone SERVAIS
avenue de l'Université 99. 1050 Bruxelles Belgique
Tel.: 02/ 48 03 70

Chaque numéro est consacré à l'étude d'un philosophe, d'un mouvement ou d'un problème philosophique, accompagnée souvent d'une bibliographie.

Vraiment internationale, la Revue recrute ses collaborateurs dans tous les pays et elle publie ses articles en allemand, anglais, espagnol, français ou italien.

Quatre numéros par an.

Abonnements: Belgique 450 FB.
Pays européens: 475 FB.
Autres pays: 500 FB.

Derniers numéros parus:

1971 N° 95-96: Hypothèse et confirmation
N° 97: L'Egalité
N° 98: Logique et Méthodologie des sciences en U.R.S.S.

Numéros prévus:
Americana
Feuerbach
Russell
Meinong
Etudes philosophiques et Informatique
Lukàcs

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by

NICHOLAS RESCHER

VOLUME 9/NUMBER 3 JULY 1972

CONTENTS

I PATRICK MAYNARD: *Vision and Convention: Recent Steps Toward a Philosophy of Pictures*
II KENTON F. MACHINA: *Vague Predicates*
III DIOGENES ALLEN: *Deliberation and the Regularity of Behavior*
IV JOHN LESLIE: *Ethically Required Existence*
V BRUCE AUNE: *On an Analytic-Synthetic Distinction*
VI JOEL J. KUPPERMAN: *Aesthetic Value*
VII ALASTAIR MCKINNON: *"Existence" in "The Existence of God"*
VIII A. AARON SNYDER: *The Paradox of Determinism*
IX HUGUES LEBLANC: *Wittgenstein and the Truth-Functionality Thesis*

Annual subscription: $18.00 (Institutions), $8.00 (Individuals).
Single Issues: $5.00 (Institutions), $3.00 (Individuals).

BASIL BLACKWELL
Oxford, England

NOTRE DAME JOURNAL OF FORMAL LOGIC

BOLESŁAW SOBOCIŃSKI, *Editor*

Managing Editor: JOHN THOMAS CANTY
Business Manager: ANN W. LAUER

Published quarterly by the University of Notre Dame in January, April, July and October of each year, the Journal is devoted primarily to logic and the foundations of mathematics (i.e. algebraic logic, classical logical systems, constructive mathematics, methodology of deductive systems, model theory, non-classical formal systems, proof theory, recursion theory, and set theory).

Articles for publication and all editorial correspondence should be sent to the *Editor, P.O. Box* 28, *Notre Dame, Indiana* 46556, *U.S.A.*

Subscriptions are $16.00 per year for institutions and $10.00 for individuals. Back volumes of the Journal are available. Orders for subscriptions and back volumes (accompanied by payment), and changes of address should be sent to the *Business Manager, P.O. Box 5, Notre Dame, Indiana* 46556, *U.S.A.*

THEORIA

A Swedish Journal of Philosophy

Edited by GÖRAN HERMERÉN *and* KRISTER SEGERBERG

CONTENTS

VOLUME XXXVII 1971 Part 1

Michael Martin: The body-mind problem and neurophysiological reduction
K. Marc-Wogau: Remarks on Nilson's paper "On the logic of historical explanation"
Herbert Hochberg: Frege on concepts as functions: a fundamental ambiguity
Lennart Åqvist: Revised foundations for imperative-epistemic and interrogative logic
David Lewis: Completeness and decidability of three logics of counterfactual conditionals
G. Germano: Incompleteness and truth definitions

THEORIA, which invites contributions in all fields of analytic philosophy, is published three times a year. Subscriptions and requests for back numbers should be sent to the publishers, Messers. C. W. K. Gleerup, Öresundsvägen 1, 222 38 Lund, Sweden. Annual subscription is Skr. 30:— (approx. 6 US $).

SYNTHESE

An International Journal for Epistemology, Methodology and Philosophy of Science

Editor-in-Chief: JAAKKO HINTIKKA
University of Helsinki and Stanford University

Review Editor: Dagfinn Føllesdal, *University of Oslo*
Associate Editor: Risto Hilpinen, *University of Jyväskylä*
Editorial Assistant: Juha Manninen, *Helsinki*

Contents of Volume 23, Nos 2/3

Roger Rosenkrantz / Inductivism and Probabilism—Keith Lehrer / Induction and Conceptual Change—Herman Vetter / Inductivism and Falsificationism Reconcilable—Lewis G. Creary / Empiricism and Rationality—Risto Hilpinen / Relational Hypotheses and Inductive Inference—Nathan Stemmer / Three Problems in Induction—Henry E. Kyburg, Jr. / Epistemological Probability.
Reviews: M. Swain (ed.) / Induction, Acceptance, and Rational Belief (Howard Smokler)—Nicholas Rescher (ed.) / Essays in Honor of Carl G. Hempel (Howard Smokler)—K. Lambert and G. G. Brittan, Jr. / An Introduction to the Philosophy of Science (R. Harré) —P. V. Tavanec (ed.) / Problems of the Logic of Scientific Knowledge (Risto Hilpinen).
Editorial Note: Arne Naess: Selected List of his Philosophical Writings in the English and German Languages. 1936-1970—Books Received.

Subscription price per volume of four issues Dfl. 90,— (US $37.38) including postage
Six issues are published yearly. Personal subscription price on request.

D. REIDEL PUBLISHING COMPANY
P.O. Box 17 / DORDRECHT-HOLLAND

MIND

A Quarterly Review of Psychology and Philosophy
edited by Professor **DAVID HAMLYN**

APRIL 1972

I.	Strong and Weak Descriptivism in Ethics	C. D. MACNIVEN
II.	Imperatives	A. BROADIE
III.	The Argument from Illusion: Objects and Objections	M. KITELEY
IV.	Illocutionary Forces and the Logic of Change	L. APOSTEL
V.	Propositions and Truth in Natural Languages	W. KNEALE
VI.	DISCUSSIONS:	
	Cohen on Evidential Support	R. G. SWINBURNE
	A Reply to Swinburne	L. J. COHEN
	'Here' and 'Now'	E. M. ZEMACH
	Reconstructing Mills "Proof" of the Principle of Utility	H. R. WEST
	Ducks, Geese and Mr. Bennett	V. PRATT
	Austin on Nowell-Smith's Conditional Analyses of 'Could Have' and 'Can'	J. C. D'ALESSIO
	Collingwood's Debt to Croce	A. DONAGAN
	Don Locke and 'Appearance-Determined Qualities':	J. KNOX
	Kirk on Quine on Bilingualism	A. HYSLOP
	Vacuosity	G. ENGLEBRETSEN
	Translation into Martian	R. HARRIS
	On the Puzzle of Self-Supporting Inductive Arguments	A. KASHER
	Benardete on Sense Perception and the *A Priori*	M. MULLICK
	'And'	P. FOULKES
VII	Critical Notice:	
	The Object of Morality. By G. J. Warnock	J. NARVESON
VIII	New Books	
IX	Notes	

Annual Subscription £1.25, post free; single copy 40p, plus postage

Published for MIND ASSOCIATION by
BASIL BLACKWELL 49 BROAD STREET OXFORD

VOLUME XIV JUNE 1972 No. 1

RATIO

edited in co-operation with
H. B. Acton (London), P. Bernays (Zürich), A. Church (Princeton), H. Cherniss (Princeton), G. Henry-Hermann (Bremen), W. C. Kneale (Oxford), E. Nagel (Columbia), K. R. Popper (London) and J. W. N. Watkins (London)
by
Stephan Körner (Bristol)

M. J. SCOTT-TAGGART	Socratic Irony and Self Deceit
OTTA WEINBERGER	The Concept of Non-Satisfaction and the Logic of Norms
N. H. PALMER	Must Clocks be Material?
MORRIS LAZEROWITZ	Moore's Ontological Program
WINFRED WEIER	Logical Problems of Epistemological Immanentism
J. MARGOLIS	Knowledge, Belief and Thought
JEROME SHAFFER	Minds, Bodies, and Theoretical Entities
REVIEWS:	
PETER WINCH	*Without Answers* (Rush Rhees)
GRETE HENRY-HERMANN	*Recht und Pflicht bei Leonard Nelson* (Christoph Westermann)

Price £1.00 net **Annual Subscription: £1.80 net**

BASIL BLACKWELL • OXFORD

INQUIRY An Interdisciplinary Journal of Philosophy and the Social Sciences
Edited by Arne Naess
and Alastair Hannay

Numbers 1-2 *Summer 1972* *Volume 15*

Editorial Statement	
Insanity and Responsibility	Herbert Fingarette
Mental Structure and Self-Consciousness	Brian O'Shaughnessy
Spinoza and Mental Health	Paul Wienpahl
Towards an Ecology of Consciousness	Frank Barron
Schizophrenia and Perception: a Critique of the Liberal Theory of Externality	James M. Glass
The Validity of Psychotherapy	B. A. Farrell
Review Discussions:	
Richard Wollheim: *Freud*	Frank Cioffi
Paul Ricoeur: *Freud and Philosophy*	Paul A. Lee
Thomas Szasz: *The Manufacture of Madness*	Ronald B. de Sousa
Roger Bastide: *The Sociology of Mental Disorder*; Michael Phillipson: *Sociological Aspects of Crime and Delinquency;* David Robinson: *The Process of Becoming Ill*	Maurice Roche
Erwin W. Strauss, Maurice Natanson, and Henri Ey: *Psychiatry and Philosophy*	A. Hannay

Published quarterly by Universitetsforlaget, P.O. Box 307
Blindern, Oslo 3, Norway

Annual Subscription: $10.00, £3·30 including postage

Volume 23 Number 3 August 1972

The British Journal for the Philosophy of Science

CAMBRIDGE UNIVERSITY PRESS

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editor: IMRE LAKATOS

Assistant Editors: COLIN HOWSON, JOHN WORRALL AND ELIE ZAHAR

Subscriptions

The *Journal* is published quarterly. The subscription price is £4·50 net in U.K. (U.S.A. $14.00) per volume post free; single issues £1·50 net in U.K. (U.S.A. $4.75) plus postage. The subscription is £3·00 (U.S.A. $9.50) for members of the British Society for the Philosophy of Science, who should order direct from the publisher. Orders may be sent to any bookseller or direct to Cambridge University Press, Bentley House, 200 Euston Road, London NW1 2DB. Orders from U.S.A. and Canada should be sent to Cambridge University Press American Branch, 32 East 57th Street, New York, N.Y. 10022, U.S.A. Second class postage paid at New York, N.Y.

Notes to Contributors

Articles, correspondence and books for review should be sent to the Editor, Professor Imre Lakatos, at the Department of Philosophy, London School of Economics, Houghton Street, Aldwych, London WC2A 2AE. Articles should not normally exceed 5,000 words, and should be accompanied by abstracts of about 100 words.

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Footnotes, which should be as few as possible, should be typed separately, double-spaced at the end of the text. They should be referred to by consecutive numbering throughout the text, in which they will be inserted in page proof. Bibliographical references should be listed alphabetically at the end. The style of references should follow current *Journal* practice.

Long quotations are printed in small type without quotes, and should be so marked in the typescript. Single quotes should be used to form the names of words and sentences, and double quotes for all other purposes. Diagrams, tables, and illustrations should be on separate sheets, with their desired position in the text clearly indicated. Symbolic formulae appearing as separate lines should be clearly distinguished in the text, and foreign letters (Greek, German, etc.) indicated in the margin. Material to be printed in italics, especially English letters used as symbols ('*A*', '*x*', etc.), should be underlined. On matters of style and layout, authors should consult CHAUNDY, T. W. *et al.* [1964]: *The Printing of Mathematics*, London: Oxford University Press.

Offprints of articles may be ordered when corrected galley proofs are returned. Fifty are supplied free ; the price of extra copies is given on the order form.

© Cambridge University Press, 1972

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the London or New York office of Cambridge University Press.

ISI Tear Service, 325 Chestnut Street, Philadelphia, Pennsylvania 19106, U.S.A. is authorised to supply single copies of separate articles for private use only.

Advertising

Particulars of advertising in the *Journal* may be obtained from Cambridge University Press, Bentley House, 200 Euston Road, London, NW1 2DB; or in U.S.A. and Canada, Cambridge University Press American Branch, 32 East 57th Street, New York, N.Y. 10022 U.S.A.

Back Volumes

Volumes 1-16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17-21 may be obtained from Cambridge University Press.

Brit. J. Phil. Sci. **23** (1972), 191–206 *Printed in Great Britain*

Numbers and Numerals

by WILLIAM C. KNEALE

1 In order to give an acceptable account of truth which avoids on the one hand the antinomy of the Liar and on the other the Tarski paradox that natural languages with pretensions to omnicompetence are all inconsistent, we must distinguish between propositions and the sentences or other propositional phrases by utterance of which propositions may be expressed (cf. Kneale [1972]). Similarly, in order to make clear what is involved in saying that predicates are true of (or satisfied by) things, and so to solve Russell's paradox in the form in which he first presented it, we must distinguish between predicates in the logical sense presupposed by such talk and the predicative phrases of grammatical analysis by utterance of which predicates may be expressed (cf. Kneale [1971]). In this paper I wish to argue that numbers belong to the same realm as propositions and predicates, and that through recognition of this fact we can escape from some notorious difficulties in the philosophy of mathematics. Briefly my thesis is that numbers are to numerals as propositions are to sentences. But without further comment this is not very enlightening, and I must therefore begin by trying to explain it in some detail.

2 In the Introduction to his [1884] Frege declared it a scandal that even the mathematicians of his day could give no sensible answer to the question 'What is the number one?' though all agreed it was a question about the simplest object of their science. More recently some philosophers have said that puzzles like this of Frege are a prime source of unnecessary bewilderment in philosophy. According to their view we should not ask for a definition of the meaning of a simple word like 'one', but rather enquire how it is used in ordinary speech, because an understanding of that use will dispel our perplexity without starting an infinite regress of pseudo-explanations. Whether or not these philosophers are right in their general conception of philosophy, they cannot be right in their treatment of Frege's question if they suppose that he started with a worry about the use of the word 'one' and similar numerals as they occur in sentences of ordinary speech like 'I have only one daughter'. For in this example the word 'one' is an adjective, and if we translated the whole sentence into German we should have to write *Ich habe nur eine Tochter*. What interested Frege was the use of 'one' and similar words in arithmetic as nouns. In his own German the point is very clear, because there the numerical noun

corresponding to our 'one' happens to have acquired the special form *Eins*. Although this is obviously derived from *eines*, a neuter adjective meaning 'one thing', it has been so specialised for the noun role in arithmetic that it is now printed with a capital letter and assigned to the feminine gender, like *Zahl*, the German word for 'number'. This, however, is a special case. In all civilised languages of which I have heard numbers are ordinarily designated either by words like 'one', 'two', 'three', which would otherwise be called numerical adjectives, or by words like 'hundred', 'thousand', 'million', which started life as numerical nouns of aggregation but have come to be used in arithmetic without regard to any things aggregated. When numerical adjectives are declinable, as many were in ancient Greek, the form used for the noun role is normally the neuter plural. Clearly adoption of these practices by our ancestors was connected with establishment of arithmetic as a discipline necessary for many practical affairs, and I have given evidence elsewhere to show that it could still produce linguistic awkwardness in Plato's time (Kneale [1962], p. 392). Because of the great influence which mathematics has in our society, we are all familiar now with the use of numerals as designations in sentences like 'Two plus two equals four' and 'Seven is a prime number', but that is not to say that we all understand perfectly what is involved in arithmetical talk. When Frege asked his question he was not working up an unnecessary fuss by repeating a fragment of common language until it came to seem mysterious because of its abstraction from the practical affairs of everyday life. On the contrary, he had noticed a linguistic innovation which seemed to be required for the development of science but was not explained properly by the learned persons responsible for it. How then should we answer his question? Was he right in thinking that there must be supersensible non-linguistic objects designated by the words 'one', 'two', 'three' when these are used as nouns in the formulation of arithmetical truths?

Already in his time there were many hard-headed persons who shrank from admitting this because they were afraid of being committed to Platonism or even to Pythagoreanism. And in order to escape from such unpopular metaphysics some who called themselves formalists thought it necessary to maintain that the numbers studied in arithmetic were no more than numerals, *i.e.* man-made signs of the sort used in counting. So far as I know, they did not dwell on the history of numerals and the comparative novelty of the substantival way in which they are used by mathematicians; but if they had been prepared to consider such evidence, they could have argued with some plausibility that when a numerical sign of adjectival origin occurs with the syntactical role of a noun it is to be understood as having what medieval logicians called material supposition, that is to say as being a quotation of its normal adjectival self. Admittedly it is no longer correct to *write* 'One is a numerical adjective of English', though the predicative phrase of this grammatical example shows clearly enough how we are to take the subject part, namely as a special sign which is not itself an adjective but something devised for referring to an adjective. In speech, however, it is still possible to make quotations without using quotation marks, and so it was perhaps not unnatural to suppose that when mathematicians talk of the number one they are merely taking advantage of an old licence in order to say of the numeral 'one' something not ordinarily accounted grammatical. Attractive as the suggestion may have seemed on first hearing a hundred years ago, it could not survive careful examination. For the realms of arithmetic and grammar are too far apart to be united. What may be asserted truly of a number by a mathematician, *e.g.* that it is prime, can have no application to a numeral, if by this is meant a pattern exhibited by certain sounds or marks used in counting. And conversely what may be asserted truly of a numeral, *e.g.* that it is monosyllabic, can have no application to a number, if by this is meant something that mathematicians of different countries can study equally well, each in his mother tongue. So now even those who pride themselves most on the hardness of their heads hesitate to identify numbers with numerals. But in retrospect we may perhaps say that the mistake of the original formalists lay in assuming an over-simplified theory of quotation.

In his Gettysburg Address Lincoln spoke of the American nation as 'conceived in liberty and dedicated to the proposition that all men are created equal'. Without altering the sense of his speech, he could have omitted the word 'that' and ended the passage by saying simply 'the proposition "All men are created equal" '. For when, immediately after the words 'the proposition', a man goes on to say 'All men are created equal', no one with a correct understanding of the situation will suppose him to refer to a sentence. It is true, of course, that the part of his speech which comes immediately after the words 'the proposition' is a specimen of a sentence; but it is clear from the context that he produces a specimen of a sentence only in order to produce a specimen of a proposition, and that he might conceivably have produced a specimen of the proposition by producing a specimen of some sentence other than that he did in fact use. Thus if the United States had been a bi-lingual country like Canada, there might have been a French version of the address in which Lincoln spoke of *la thèse 'Tous les hommes sont créés égaux'*. Similarly when we who speak English talk of the number *one* we must produce a specimen of the English numeral *one*, but it is clear from the context that we are not interested in the numeral for its own sake and that, although talk of the number involves a kind of quotation (which I have indicated by putting italics after the word 'number' as well as after the word 'numeral'), there can be no objection to translation of what we quote in a mathematical as distinct from a grammatical statement, because we are concerned only with its semantic role, which is something it shares or may share with signs of different sound or shape.

If I am right, the use of the word 'one' which is involved in arithmetical talk of the number *one* is something like the use of a general term with supposition of the kind called simple by medieval logicians. Their standard example of such use was the old and much debated assertion *Homo est species* ('Man is a species') in which *homo* was said by them to stand for its own sense (*supponere pro suo significato*) (cf. Kneale [1962], pp. 252 ff.). Presumably they talked of simple supposition here because the supposition, or reference, of the subject term seemed to them to coincide with its signification, from which supposition of any other kind would have to be distinguished. Ockham and other nominalists tried to include it under the heading of material supposition; but for this purpose they had to say that the old example was a statement about the *intentio*, or mental sign, which they assumed to precede the spoken word *homo* in the logical order of things, much as that may be said to precede the written word (*ibid.*, p. 269). Another analogy, which may perhaps be thought more illuminating, can be found in Frege's use of phrases such as *der Begriff 'Mensch'* ('the concept "man" '). In his paper *Begriff und Gegenstand* he put the word *Mensch* in italics instead of ordinary quotation marks but it is clear that his intention was always to indicate a kind of quotation. When writing most carefully he would sometimes explain that a single word like *Mensch* occurring in quotation marks after the word *Begriff* was short for a predicative phrase like *ist ein Mensch.* But to anyone familiar with his philosophy of language it is obvious that he never thought of a concept as being itself either a word or a string of words.

Unfortunately neither analogy is perfect for my purpose. When medieval logicians used the word *homo* in the special way I have indicated they did not distinguish its significate, or sense, from the form humanity. And similarly in Frege's philosophy of language there is no distinction between the concept *man* and the property humanity. What I want to suggest is that the propositions, predicates, and numbers expressed by various token utterances are kinds to which those utterances belong in virtue of their roles in communication, and that, although they are not to be identified with sentences, predicative phrases, and numerals respectively, since these latter are kinds to which utterances belong in virtue of their sensible appearances, nevertheless they too may be indicated by production of specimens. Thus if anyone wishes to refer to a predicate, his simplest way of designating it is to say the words 'the predicate' and then produce a specimen of the predicate that interests him. Similarly, I hold, if anyone wishes to refer to a number, his simplest way of designating it is to say the words 'the number' and then produce a specimen of the number that interests him. Inevitably the specimen will also be a token of a numeral of the speaker's language, but when he introduces it by saying first 'the number', he shows clearly what he intends it to exemplify, and does so in just the same way as he shows what interests him when he says 'The sound "bang" is explosive'. In either case the device we call quotation is production of an example for the purpose of specifying a kind within a certain category.

3 What, then, is the role in communication of all the utterances in many different languages which I wish to class together as specimens of the number one? In order to answer this and all the similar questions that may be asked with regard to numbers other than one, we must consider first how we use the word 'number', and in particular how we use it outside arithmetic, since we are not likely to get much philosophical light from considering only its use inside arithmetic, where everything we say presupposes familiarity with the notion. Now the commonest use outside arithmetic is in statements such as 'The number of the apostles is twelve', where we give what may also be called the total of a class. If, for example, it is known that the number of the apostles is twelve and that for each apostle the number of his fingers is ten, we can easily conclude in accordance with the multiplication table that the number of the fingers of all the apostles taken together is 120. But to say that the number of the apostles is twelve is only to say in a rather special way that there are just twelve apostles; and the more elaborate form of expression seems to have been introduced at a comparatively late stage in the evolution of language for the express purpose of putting data in the way most convenient for application of arithmetical techniques. In order, then, to make further progress we must consider what is involved in saying such things as that there are just twelve apostles, and here it will be convenient to use a few printed symbols of formal logic for the purpose of focusing attention on what is essential.

Starting with any predicative phrase '$F(x)$', we can build up a sequence of propositional phrases of ever increasing complexity by use of the existential quantifier, the sign of conjunction, the sign of difference (or non-identity), and an unlimited supply of variables. If we take 'x', 'y', and 'z' as the first three variables, the sequence starts as follows:

$$\begin{array}{ll}\exists x\,.\,F(x) & \\ \exists y\exists x\,.\,F(x)\ \&\ F(y)\ \&\ x \neq y & (\alpha)\\ \exists z\exists y\exists x\,.\,F(x)\ \&\ F(y)\ \&\ x \neq y\ \&\ F(z)\ \&\ x \neq z\ \&\ y \neq z & \end{array}$$

For convenience of comparison I have indicated the scope of generalisation in each formula by means of a single dot and then disposed the parts of successive formulae in such a way as to show clearly how each new formula is built by additions to its predecessor. The rule of construction is to add on the left another existential quantifier with a new variable while adding on the right after conjunction signs a phrase of the form '$F(\ )$', with the same new variable inside the bracket, and as many new difference clauses as can be made with previously used variables in the first place and the new variable in the second place. In ordinary language, where we have the token reflexive device at our disposal, we can produce the same results more simply by saying 'There is an F thing, and another, and another, . . .'. But in the symbolism of formal logic we must start with the more complicated formulae I have just set out and then, if we wish, replace them by the shorter formulae:

$$\begin{array}{cl}\exists x\,.\,F(x) & \\ \exists\exists x\,.\,F(x) & (\beta)\\ \exists\exists\exists x\,.\,F(x) & \end{array}$$

Obviously such abbreviation is quite safe, since there can be no doubt about the way in which to return from the shorter formulae to the originals. And the result is even better than that achieved in ordinary language by use of the phrase 'and another', since it involves repetition of the existence sign in such a way as to make numerals of a basic type like the Roman 'I', 'II', 'III'.

There is, however, one defect in our notation. Like that of the Romans, it contains no zero sign. In order to make good this shortcoming let us adopt two more conventions: (1) that whenever one or more tokens of the existential sign '$\exists$' occur before a variable the whole string is to be enclosed within angular brackets, and (2) that '$\langle\ \rangle x\,.\,F(x)$' may be used as an abbreviation for '$\sim\exists x\,.\,F(x)$'. The result is, of course, that an assertion with the sense of 'Nothing is F' can be expressed by *omission* of the sign '$\exists$' from the formula corresponding to 'Something is F'. And so we are able to construct the sequence of formulae:

$$\begin{array}{cl}\langle\ \rangle x\,.\,F(x) & \\ \langle\exists\rangle x\,.\,F(x) & \\ \langle\exists\exists\rangle x\,.\,F(x) & (\gamma)\\ \langle\exists\exists\exists\rangle x\,.\,F(x) & \end{array}$$

in which the first operator is simply a pair of empty brackets and each succeeding operator contains one more occurrence of '$\exists$' than its predecessor.

Since the formulae of sequence (γ) can be abbreviated very easily to

$$\begin{aligned} &0x \,.\, F(x) \\ &1x \,.\, F(x) \\ &2x \,.\, F(x) \\ &3x \,.\, F(x) \end{aligned} \qquad (\delta)$$

and these in turn can be read as variants of the English phrases

No thing is F
One thing is F
Two things are F (ϵ)
Three things are F

it is clear that the adjectival numerals of ordinary language have a semantic status like that of the existential quantifier and may indeed be described as signs of existence in all its various complications. So far as the numeral 'one' is concerned, this conclusion is not at all surprising, since we are accustomed to thinking of the word as a variant of the indefinite article which has been recognised as a symbol for asserting existence in sentences like 'A man has landed on the moon'. And if its application to numerals other than 'one' seems less obvious, the reason may perhaps be simply that our intuition is dimmed by an unfortunate tradition of talking in logic about quantity and quantifiers where it would be more appropriate to talk of quotity and quotifiers. Possibly the trouble started with an ambiguity of the Greek word *posotēs*.

Of the formulae in sequence (γ) all but the first are compatible with their successors. Thus the second has the sense of 'One thing at least is F'. When, however, anyone says that the number of the apostles is twelve, he means not merely that there are at least twelve apostles but that there are just twelve apostles, *i.e.* twelve and no more. If we find it convenient to do so, we can introduce special operators with the sense required for making such statements succinctly. In the table printed below they are given on the left opposite their equivalents in the notation we have considered so far:

$0x \,.\, F(x)$	$\sim 1x \,.\, F(x)$
$^*1x \,.\, F(x)$	$1x \,.\, F(x)\ \&\ \sim 2x \,.\, F(x)$
$^*2x \,.\, F(x)$	$2x \,.\, F(x)\ \&\ \sim 3x \,.\, F(x)$
$^*3x \,.\, F(x)$	$3x \,.\, F(x)\ \&\ \sim 4x \,.\, F(x)$

Because use of the zero sign is already incompatible with use of any other numeral in the same formula, it requires no prefix with the sense of 'just', but in each of the other cases addition of the asterisk is required to make what is sometimes called an exact numeral.

In all this there is nothing much new except the suggestion that adjectival numerals are themselves signs of existence or non-existence, as distinct from indices that may be attached to such signs. Certainly Frege was well aware of the connection between numbers and existence, since he wrote in one famous passage 'Affirmation of existence is . . . nothing but denial of the number nought' (Frege [1884], §53). But instead of mentioning it in order to explain the notion of number, which was his main concern, he wrote as though he thought reference to a number by means of a numerical noun was more explicit and fundamental than use of an adjectival numeral or the verb 'exist'. In one place he declares 'If I say "The king's carriage is drawn by four horses", then I assign the number four to the concept "horses that draw the king's carriage" ' (*ibid.*, §46). And a little later he explains his thought more fully by saying 'The construction in "four thoroughbred horses" fosters the illusion that "four" modifies the concept "thoroughbred horses" in just the same way as "thoroughbred" modifies the concept "horse" ' (*ibid.*, §52). He was right, of course, in pointing out a difference between numerical adjectives and others, but he failed to see that their peculiarity becomes intelligible when they are treated as logical operators like those he himself introduced in his theory of general statements, and he made a mistake when he tried to eliminate them in favour of numerical nouns. Once this has been recognised, it is natural to reverse his order of explanation by exhibiting the numerical nouns of ordinary language as quotations of the numerical adjectives of ordinary language.

Having introduced the signs '0', '1', '2', *etc.* as operators, we cannot without notice or explanation go on immediately to use them alone as designations. And although use of them as designations involves a kind of quotation, it may perhaps be misleading to print them in inverted commas when our purpose is to refer to numbers. I propose, therefore, to write '#0', '#1', '#2', *etc.* in contexts where it would be proper to speak of the number 0, the number 1, the number 2, *etc.* I realise, of course, that it would be extremely tedious to put the sign '#' before every occurrence of a numeral with the simple supposition of which medieval logicians wrote, and I do not suggest that it is necessary to maintain this standard of rigour at all times; but I think it important to make clear to ourselves occasionally that '2+2 = 4' is short for '#2 + #2 = #4'.

In my [1962] and some more recent writings I have used '§[P]' for 'the proposition that P' and '§x[$F(x)$]' for 'the predicate being-F'. In the same way, using 'N' as a schematic letter to mark a place for adjectival numerals, we may write

$$\#N = \S f[Nx.f(x)].$$

By drawing attention to a kinship between numbers and predicates this notation makes it plausible to say of numbers in general what Frege said of simple existence, or the number one, namely that they are concepts of the second order. But some justification is required for use of the same notation in the three separate cases I have mentioned, and the only way to provide this is to give an interpretation that holds throughout. For this purpose we may take the notion of truth as fundamental and explain the three uses of '§' by the following equivalences:

$\S[P]$	That which is true if and only if by the same token P.
$\S x[F(x)]$	That which is true for a thing, say x, if and only if by the same token $F(x)$.
$\S f[Nx\,.\,f(x)]$	That which is true for a predicate, say $\S x[f(x)]$, if and only if by the same token $Nx\,.\,f(x)$.

In each case the sign '§' may be called an intensional abstractor, or nominaliser; but any complex sign which it serves to form is a definite description that fails to designate anything if that on which the nominaliser operates, *i.e.* the expression contained in square brackets, does not express a proposition, a predicate, or a number, as the case may be.

Although we now have the sign '#' available as a nominalising prefix that can be put before an adjectival numeral and read as equivalent to the phrase 'the number', we have still to explain the construction of the English phrase 'the number of things which are F'. In Frege's terminology a number described by such a phrase is called an *Anzahl*, or total number, and it is important for his purposes that the notion of a total number should be definable without reference to the abstract numbers (*Zahlen*) of which we speak in pure arithmetic. For each of the latter is to be defined as the class of all classes which have the same total number as a given class from a certain standard progression of classes. Since it is certainly permissible to say, as he does, that a total number is something characteristic of a class, let us provisionally write '$\mathrm{tn}\{x\,.\,F(x)\}$' for 'the total number of things which are F'. But instead of trying to define the total number of a class, as he does, by reference to what the class shares with all those classes with which it can be put into one-to-one correspondence, let us consider how in practice we establish that the total number of F things is a certain natural number N. Clearly we rely on the principle

$$\mathrm{tn}\{x\,.\,F(x)\} = \#N \quad // \quad {}^{*}Nx\,.\,F(x),$$

where ‘//’ has the sense of ‘if and only if by the same token’. And in the light of this principle it seems natural to define the number of things that are F by the equivalence

$$\mathrm{tn}\{x\,.\,F(x)\} = \text{℩}y\,\exists n[y = \#n\ \&\ {}^{*}nx\,.\,F(x)],$$

or more simply by the equivalence

$$\mathrm{tn}\{x\,.\,F(x)\} = \#n[{}^{*}nx\,.\,F(x)],$$

in which the sign ‘#’ is allowed to take over the full duty of a prefix introducing a definite description. In ordinary English supplemented by the variable ‘n’ the second suggestion can be expressed by saying ‘The number of things which are F is the number n such that just n things are F’. Such a definition depends of course on the possibility of using a numerical variable as an adjective or operator, but as might be expected such use is a commonplace in mathematical texts. Hardy and Wright ([1938], p. 14) for example, when talking of the Fermat numbers defined by the formula

$$F_n = 2^{2^n}+1,$$

say ‘If F_n is a prime p, then a regular polygon of p sides can be inscribed in a circle by Euclidean methods’ and later on the same page ‘There are at least n odd primes not exceeding F_n’. In short, for those who recognise the possibility of using ordinary adjectival numerals with simple supposition the notion of a total number raises no more difficulty than that of a number *tout court*. All talk of classes is superfluous here, though undoubtedly permissible.

4 Since numbers are kinds to which utterances may belong by virtue of their role in communication, it is easy to see that our ability to generalise about them depends on our understanding of what we do in using numerals for counting. But it would be misleading to express this insight by saying that the natural numbers are free creations of the human mind. For what interests us when we call an utterance a specimen, or expression, of a number is its relation to something not made by men, and all we can say about numbers in mathematics derives from this relation.

Just as an utterance cannot be a specimen of a proposition without presenting a possible state of affairs, or a specimen of a predicate without presenting a property that things may have, so also an utterance cannot be a specimen of a number without presenting a possible *quotity*, or extent of exemplification. For convenience of review the various notions involved in this comparison may be set out as follows:

	Sign Types	*Presentations*	*Realities Presented*
1.	Propositional Phrases (*e.g.* sentences)	Propositions	States of Affairs (*e.g.* facts)
2.	Predicative Phrases (Predicates in the grammatical sense)	Predications (Predicates in the logical sense)	Properties (Attributes)
3.	Quotifiers (Numerals)	Quotifications (Numbers)	Quotities (Extents of exemplification)

I have taken the word 'presentation' from William of Shyreswood, who speaks of *praesentatio alicuius formae ad intellectum* when he needs a term to cover the work of signification done by different parts of speech (cf. Kneale [1962], p. 246). It is metaphorical, of course, but not likely to be dangerous on that account, and for my purpose it has the great merit of being free from any special link with expression, which is the relation of an utterance to a linguistic kind listed in my middle column, or designation, which is indeed a relation of an utterance to an independent reality, but not one that can be taken as fundamental in semantics.

In his theory of language Frege made no attempt to distinguish presentations from realities presented, *e.g.* concepts from properties; and when he came to consider numbers the account he gave of them was appropriate rather to quotities. Perhaps he was unwilling to attach any importance to the role of adjectival numerals in ordinary language because he thought that reflection on such matters could lead only to a mistaken identification of numbers with numerals. And it is clear that, once he had decided to ignore the way in which quotities are presented by means of adjectival numerals, his only hope of providing a foundation for arithmetic was to consider the common features of classes that are all alike describable by some common noun such as 'pair', 'dozen', 'score'. But he had also another motive for trying to deal with the notion of quotity afresh by concentrating attention on one-to-one correlation between classes, namely a declared wish to define an *Anzahl*, or total number, in such a way that it might be identified with what Cantor called the *Mächtigkeit*, or power, of a set. For with such a definition he thought that the natural numbers might be exhibited as finite cardinals, suitable no doubt for study in infant schools but nevertheless numbers in the same sense as Cantor's transfinite cardinals. Against this project it may perhaps be objected that we have no reason for regarding infinity as a quotity (*i.e.* for admitting 'Infinitely many' as a possible answer to the question 'How many?') until we already have free use of the concept of natural number and can say that the objects falling under it form a sequence without end. For from this it seems to follow that talk of cardinals, whether finite or transfinite, belongs to a different level of language from that in which we first mention natural numbers. But we have no reason to deny that even finite quotities may be indicated without recourse to adjectival numerals, and what Frege says of the natural numbers seems in fact to fit very well if taken as an independent account of the quotities normally presented by specimens of the natural numbers. In the language of later set theory it may be expressed simply by the statement that each finite cardinal is the cardinal of a finite ordinal from the sequence

$$\Lambda,\ \{\Lambda\},\ \{\Lambda, \{\Lambda\}\},\ \{\Lambda, \{\Lambda\}, \{\Lambda, \{\Lambda\}\}\}, \ldots$$

Obviously the ordinals listed by Frege as paradigmatic members of his finite cardinals resemble the natural numbers

$$\#\langle\ \rangle,\ \#\langle\exists\rangle,\ \#\langle\exists\exists\rangle,\ \#\langle\exists\exists\exists\rangle, \ldots$$

in that they form a progression, where every item except the first is immediate successor to just one other item and every item without exception has just one immediate successor; but he did not suggest that they themselves might be called natural numbers, and I believe that, if asked for his opinion of such a proposal, he would have rejected it as absurd. It is interesting therefore to notice a more drastic simplification which had been put forward already in his time by Dedekind and still has supporters.

5 Having defined a system of things as simply infinite if it satisfied four conditions that would now be said to characterise a progression, Dedekind wrote ([1888], §6, 73):

> 'If in the consideration of a simply infinite system N which is ordered by a projection ϕ we disregard entirely the special nature of the elements, retain merely their distinguishability, and conceive only the relations in which they are set one to another through the ordering projection ϕ, then these elements are called ***natural numbers*** or ***ordinal numbers*** or simply ***numbers***, and the basic element 1 is called the ***basic number*** of the ***number series***. Having regard to this abstraction or freeing of the elements from every other content, we may rightly describe the numbers as free creations of the human mind. The relations or laws which are derived solely from the conditions α, β, γ, δ and are therefore always the same in all ordered simply infinite systems constitute the immediate object of the *science of numbers* or *arithmetic*.'

When Dedekind says that numbers are a free creation of the human mind, he seems to mean primarily that they are abstractions; but I think he may also wish to suggest that the notion of an infinite sequence which he uses in his definition of numbers is accessible to us only because we ourselves can construct such sequences—or at any rate lay down rules for specification of their members *ad libitum.* That the numbers do in fact form a progression satisfying the conditions of Dedekind, or the very similar conditions of Peano, no one will deny; but it is paradoxical to identify numberhood with membership of a progression, whether man-made or not, because anyone who can talk of the progression of natural numbers is in a position to indicate infinitely many others that do not contain all the natural numbers.

Given the progression of natural numbers 0, 1, 2, 3, . . ., which begins with zero in accordance with Peano's final version of his theory, selection of every second item is enough to produce the progression 1, 3, 5, 7, . . ., which satisfies Peano's axioms equally well, in the sense of providing an equally acceptable, though unintended, interpretation of all his extra-logical signs. And on this basis it is possible to erect a system of odd arithmetic that contains analogues for all the statements of ordinary arithmetic derivable from the standard interpretation of Peano's axioms. Since the origin and the successor relation of this system are both different from those of ordinary arithmetic, so too, of course, are the operations of addition and multiplication we introduce by the recursive definitions

$$\begin{cases} x \oplus 1 = x \\ x \oplus n' = (x \oplus n)' \end{cases} \quad \begin{cases} x \otimes 1 = 1 \\ x \otimes n' = (x \otimes n) \oplus x \end{cases}$$

If, however, we distinguish them, as I have just done, by special signs, we can provide for their translation into the symbolism of ordinary arithmetic by the rules

$$x \oplus y = (x+y)-1$$

$$x \otimes y = \frac{x \times y-(x+y)+3}{2}$$

and then verify without difficulty that

$$3 \oplus 5 = 7 \qquad 5 \oplus 5 = 9 \qquad 5 \oplus 7 = 11$$

$$3 \otimes 5 = 5 \qquad 5 \otimes 5 = 9 \qquad 5 \otimes 7 = 13$$

There is nothing in these novelties that need surprise us and certainly nothing that should make us doubt the propriety of treating odd arithmetic as a genuine interpretation of formal arithmetic, alternative to that provided in ordinary arithmetic. For the systematic correspondence between odd arithmetic and ordinary arithmetic is confirmed by the fact that the equations I have just listed as examples of odd arithmetic can also be obtained from ordinary arithmetic by putting '⊕' for '+', '⊗' for '×', '1' for '0', '3' for '1', '5' for '2', and so on, though the equations of odd arithmetic are certainly not translations of the equations of ordinary arithmetic from which they are obtainable in this way by substitutions.

But elaboration of odd arithmetic as an alternative interpretation of formal arithmetic involves rejection of the thesis that numberhood is simply membership of a progression. For a progression whose members are just the odd numbers cannot by any feat of abstraction be so presented that it appears to contain all the natural numbers as its members.

Russell, who was familiar with such objections to Dedekind's theory, argued ([1919], pp. 7 ff.) that there could be no satisfactory foundation for arithmetic without a definition of the concept of a total number because it was only through this that the signs of arithmetic could get their usual sense and application. And here he was right, though he went astray when he followed Frege in trying to define the notion of a total number through that of one-to-one correlation between classes. The concept of a total number is essential for arithmetic simply because understanding of it provides the indispensable link between the adjectival and the substantival use of numerals. To talk of the total number of F things is to talk of the number n such that just n things are F, and anyone who is in a position to use such phraseology is in a position to go on to the elaboration of arithmetic. In this task he may use Peano's axioms as premisses if he finds that convenient, since they can be seen to hold of numbers and their relations, but he is certainly not confined to thinking of numbers as items in some progression or other. Nor could he develop arithmetic as it has been developed, if he were content to think of it as no more than the general theory of progressions, unconnected with the quotifiers of the language in which it is formulated. For arithmetic, in the large sense in which it is identical with the theory of numbers, involves theorems that are most naturally expressed by combinations of the adjectival and the substantival idioms, as in the statement quoted above that there are at least n odd primes not exceeding $2^{2^n}+1$.

In recent writing ([1969], p. 44) Quine has tried to answer Russell by saying 'We can define "there are n so-and-sos" without ever deciding what numbers are apart from their fulfilment of arithmetic. That there are n so-and-sos can be explained simply as meaning that the so-and-sos are in one-to-one correspondence with the numbers up to n.' When he speaks here of arithmetic, he means, of course, formal arithmetic in which we derive consequences from the axioms of Peano without consideration of the meanings of any of the extra-logical words used for expression of them, and in particular without thinking of Peano's numbers as anything but items in a progression. But it is rather surprising that he makes no comment on his own use of the variable 'n' in the common mathematical way to mark a gap that may be filled with an adjectival numeral of the basic type as well as to mark a gap for which the proper filling is a substantival numeral. For he has just announced his intention to reply to what Russell said long ago about the notion of an *Anzahl*, and in this context his acceptance of the adjectival idiom is significant because it reveals his unofficial opinion that talk about the total number of so-and-sos is equivalent to talk about the number *n* such that there are just *n* so-and-sos. If questioned, however, he would presumably declare that the adjectival idiom is a loose way of speaking which he could avoid by saying 'That the total number of so-and-sos is *n* can be explained simply as meaning that the so-and-sos are in one correspondence with the numbers up to *n*.'

In a later paragraph (*ibid.*, p. 45) Quine emphasises that distinct domains of objects yield distinct models and admits that for this reason it is not right to say, as is sometimes said, that numbers are any things fulfilling formal arithmetic. But he concludes 'The subtle point is that any progression will serve as a version of number so long and and only so long as one sticks to one and the same progression. Arithmetic in this sense is all there is to number; there is no saying absolutely what numbers are; there is only arithmetic.' When we take this passage together with his earlier remarks about definition of the notion of *Anzahl*, it appears that what he calls a version of number must really be a system of numerals with which it is possible to count things and so to determine their total number. For it is certainly not the case that we human beings, acting either as accountants or as pure mathematicians, have before us a set of equally qualified progressions from among which we must appoint just one to the useful and honourable post of natural number sequence. But it is true that we cannot express numbers (*i.e.* possible quotifications) without establishing a system of adjectival numerals (*i.e.* quotifiers) which is itself a progression in as much as each of its members except the zero sign is successor to a single numeral and each of its members without exception has a single numeral for its successor. And it is also true that when we designate numbers, as we do in arithmetic, for the purpose of talking about them we must use some nominalising device in connection with our adjectival numerals. What makes Quine's theory counter-intuitive is his assumption that numerical *expressions* must be introduced by reference to numerical *designations*. For this not only reverses the proper relation of the two semantical notions but commits him to a denial of the obvious fact that numerical *expressions* are already available in the logical symbolism needed for formulation of Peano's axioms.

REFERENCES

Dedekind, J. W. R. [1888]: *Was sind und was sollen die Zahlen?*
Frege, G. [1884]: *Die Grundlagen der Arithmetik.*
Hardy, G. H. and Wright, E. M. [1938]: *The Theory of Numbers.*

Brit. J. Phil. Sci. 23 (1972), 207–245 *Printed in Great Britain*

Situational Determinism in Economics*

by SPIRO J. LATSIS

Introduction.
1 *Perfect and Monopolistic Competition: Two Variations on the Neoclassical Programme of Situational Determinism.*
  (*a*) *Perfect Competition and Monopoly.*
  (*b*) *Monopolistic Competition.*
  (*c*) *Duopoly and Oligopoly: Two Apparent Exceptions.*
2 *Stagnation of the Neoclassical Programme of Situational Determinism and Emergence of Rivals.*
  (*a*) *The Appraisal of Perfect and Monopolistic Competition.*
  (*b*) *Popper's Rationality Principle.*
  (*c*) *The Emergence of Economic Behaviouralism.*
3 *An Appendix on Methodological Controversies: Falsificationism, Conventionalism and Research Programmes in Economics.*
  (*a*) *Neoclassical Apologetics versus Falsificationism.*
  (*b*) *A Falsificationist Plea for Metaphysical Statements in Economics.*
  (*c*) *The 'F-twist' versus Chamberlinian 'Realism'.*

INTRODUCTION

This paper attempts to identify and appraise two rival research programmes in a major branch of modern microeconomic theory: in the theory of the firm. I shall call them 'situational determinism' and 'economic behaviouralism'.

In section 1 I argue that the much publicised so-called Chicago-Chamberlin controversy[1] did not address itself to fundamentals but was primarily a family quarrel between two slightly differing branches of the same dominant research programme, namely, 'situational determinism'. The programme is identified and its main applications in the theory of the firm are discussed. In section 2 I suggest that all major variants or models of the basic underlying research programme of '*situational determinism*' are

---

* The author wishes to acknowledge helpful criticisms by P. T. Bauer, L. P. Foldes, L. Hurwicz, T. W. Hutchison, K. Klappholz, I. Lakatos, J. W. N. Watkins and B. S. Yamey; although, I take it, none of them will agree with everything in the present version.

[1] For interesting and informative presentations of this controversy see Stigler [1949] and Archibald [1961] and the references given there.

Received 25 *April* 1971

defective. I also examine the philosophical background of situational determinism and of its most characteristic component, the 'Rationality Principle', as formulated by Popper and I sketch a budding rival programme. Finally, in section 3 I critically examine some influential schools of thought in the methodology of economics from the point of view of the methodology of scientific research programmes.

1 PERFECT AND MONOPOLISTIC COMPETITION: TWO VARIATIONS ON THE NEOCLASSICAL PROGRAMME OF SITUATIONAL DETERMINISM

A detailed comparison of the theories of perfect and monopolistic competition reveals that their rivalry is superficial and that they form parts of the same dominant research programme with one identifiable 'hard core', one 'protective belt' and one 'positive heuristic'.[1] For reasons to be presented below, this research programme may be conveniently termed the *neoclassical programme of situational determinism.*

It will be argued that the dominant approach to the explanation of business behaviour from Adam Smith[2] through Marshall to Chamberlin has been inspired and regulated by situational determinism.[3]

Some very general, regulative, philosophical principles characterise this programme:

1. The approach is individualistic: phenomena of market behaviour are explained in terms of individual human agents acting in a social situation.

---

[1] For these concepts cf. Lakatos [1968*b*], [1970] and [1971*a*].

[2] The main thrust of the theory of perfect competition derives from the idea that because of the mass action of buyers and sellers each individual seller faces a given price and produces a predetermined output.

The assumptions proposed by Adam Smith have been transferred almost unchanged to the neoclassical programme. Smith assumed:

(1) That there is a very large number of sellers in the market.
(2) That each seller takes decisions independently of every other.
(3) That there is free entry of new sellers in the market and free exit of existing sellers.
(4) That buyers and sellers have a tolerable knowledge of market opportunities.

(Cf. A. Smith [1776], chapter VII and chapter X, part 1. For a fuller discussion of the emergence of perfect competition theory see Stigler [1957].)

Adam Smith was not entirely consistent in his adherence to these assumptions, and at times explicitly noted the presence or even the prevalence of concerted action both among sellers and buyers. (cf. A. Smith [1776], Chapter VII. However, these assumptions represent the *main* strand of Smith's thought; see, *e.g.* Stigler [1957].)

Subsequently, this programme developed with Cournot's introduction of the infinitesimal calculus (Cournot [1838]) and Marshall's detailed discussion of the model. Marshall also discussed its limitations (Marshall [1890]) such as 'decreasing costs'. Marshall's strongest doubts concerning the theory of the firm are voiced in his [1919] but there are also traces in his *Principles of Economics* ([1890]). See pp. 374–5 and p. 458, footnote 1.

[3] It may be argued, although this will not be undertaken in the present paper, that the neoclassical theory of consumer behaviour has espoused a very similar, if not identical, programme.

2. The rational choices of the individual agents are so constrained by their situation that only minimal psychological assumptions are required to explain their actions.
3. Behaviour is animated by the principle that rational agents act appropriately to the 'logic of the situation'. Popper calls this principle the *Rationality Principle.*

Situational determinism should not be confused with physical determinism. If somebody's hand comes into contact with a very hot substance it will (physical circumstances permitting) be promptly withdrawn from the heat source. The explanation of this type of behaviour is largely physiological. This physiological aspect of behaviour, though indispensable, is not a central factor in explaining the types of behaviour investigated by economists employing situational determinism. Rather than appeal to physiological constraints, situational determinists invoke the constraining nature of the decision making agent's (appraisal of his) situation.

The application of this approach to the theory of the firm resulted, after a series of refinements, in the following set of assumptions which constitute the 'hard core' of the neoclassical theory of the firm:

(*i*) *Profit maximisation.*
(*ii*) *Perfect knowledge.*
(*iii*) *Independence of decisions.*
(*iv*) *Perfect market.*

'*Profit maximisation*' refers to the seller's motive to maximise the difference between total cost and total revenue. '*Perfect knowledge*' and '*perfect market*' are required to render determinate the results of profit maximisation. The '*independence*' clause is necessary for a competitive market.

(*a*) *Perfect Competition and Monopoly.*

The simplest exemplification of situational determinism is the perfectly competitive situation. The perfectly competitive model is characterised by the hard core postulates (*i*)-(*iv*) as well as by several auxiliary assumptions, namely, (*v*) product homogeneity, (*vi*) large numbers and (*vii*) free entry and exit.

Under the conditions characterising perfect competition the decision maker's discretion in choosing among alternative courses of action is reduced simply to whether or not to remain in business. The decisions are seen as a strictly determinate process. The price which the firm can secure is a datum and so are consumer preferences and technological conditions. In equilibrium, average costs of all producers are at their minimum, and there are no profits beyond the minimum rewards required to enable the entrepreneur to continue operations. Thus, in this approach, such activities of the entrepreneur as decisions on prices, the search for information, the organisation of decision making, the choice of method of production and appeal to buyers and also the psychology of decision making, are all taken as given. The neoclassical approach may perhaps be fairly termed as envisaging entrepreneurs without entrepreneurial functions or, to put it another way, *decision makers without decision procedures.*[1] In this programme the internal structure and characteristics of the decision making unit constitute merely irrelevant noise.[2] Any disturbance to the system triggers off negative feedback which restores its initial state of 'zero profit' equilibrium. Moreover, the mere fact that the firm survives in the perfectly competitive industry is proof that its behaviour is—relatively[3]—optimal, since it would otherwise have been eliminated by a 'natural selection' procedure.

Behaviour under perfectly competitive conditions is assumed to be motivated by the overriding goal of the decision-makers to maximise profits. As we have already indicated, this motivational assumption features in the hard core of all neoclassical models. Indeed, under perfectly competitive conditions, profit maximisation becomes plausible and inconspicuous. In this scheme the decision maker's freedom of choice is spurious: it is confined to two alternatives: produce that output which will just keep the firm in business or go bankrupt. Under perfect competition entrepreneurs do not really *compete* with *each other*. The situation may be compared to that of a player in an *n*-person game where *n* is very large. Such games are reducible to one-person games against nature where the opponent has no objectives and no known strategy. The 'nature' of perfect competition is unusually strict in allowing a choice of either following a single strategy or going under.

Thus in the perfectly competitive model the assumption of profit maximisation is not central. To say that a seller under perfect competition deliberately chooses a course of action to maximise profits is analogous to saying that a member of the audience is maximising if he runs out of the single exit available to him in a burning cinema.[1] I shall call situations, where the obvious course of action (for a wide range of conceptions of rational behaviour) is determined uniquely by objective conditions (cost, demand, technology, numbers, *etc.*), *'single exit' or 'straightjacket' situations.*

---

[1] Cf. *below*, pp. 222–229.

[2] Higgins, in his [1939], struggling with the anomalies arising from indeterminate solutions, suggests that if we further specified the entrepreneurs' motivation, unambiguous equilibrium solutions could be found even in cases of indeterminacy. Rothchild [1947] also recognises that the dependence of neoclassical analysis on environmental factors is one sided. Enke, in his [1951], argues that Chamberlin's monopolistic competition theory (which according to Enke is a long-run theory of aggregate competitive behaviour), is situationally determinate while Mrs Robinson's 'theory of monopoly' is not.

[3] According to the weakened view (in the sense that the perfect knowledge assumption is relaxed) of neoclassical optimisation theory expounded by Machlup, it is subjective costs which are minimised and equalised with marginal revenue. It is, therefore, only in relation to other producers that a perfect competitor's behaviour is optimal. Cf. Machlup [1946].

Pure monopoly, usually regarded as the exact opposite of perfect competition is in fact its heuristic twin. Assumptions (*i*), (*ii*), (*iii*) and (*iv*)[2] above remain intact with the difference that (*iii*) no longer relies on the multitude of homogeneous sellers but on the uniqueness and unthreatened status of the monopolist who maximises on the basis of his knowledge of the market conditions and the application of the simple optimising rule. As with perfect competition, so with monopoly, the 'rational' decision maker will arrive at the uniquely determined optimal decision by a simple calculation. (In the case of perfect competition the firm ceases to increase its output when one extra unit, if produced, would cause costs to exceed the going price. Under pure monopoly, the firm will cease increasing its output when the cost of producing an extra unit would exceed the revenue to be gained by lowering price just enough to sell this extra unit.)

'Single-exit' methods of analysis have been extensively applied to business behaviour. The temptations for the social scientist to employ such a method are considerable:

1. *It makes the application of mathematical techniques easier.*
2. *It renders his field relatively autonomous (and therefore immune to criticism from the outside) by justifying his exclusive concentration on the description of the purely objective economic situation.*
3. *Situational descriptions are more amenable to empirical test than are descriptions of states of mind and behavioural generalisations.*

Situational determinism assumes that the decision maker knows his situation better than he is assumed to know it in the conditions usually investigated in decision theory. It is part and parcel of the neoclassical programme to specify situations which uniquely determine behaviour. Calculation of the uniquely appropriate behaviour is reduced to elementary logic and trivial common sense. We concentrate on the *logic* of the agent's *situation* and are spared the complexities of the *psychology of the agent* in that situation. The motivational and evaluational characteristics of the acting agent become largely ornamental.

Having outlined the hard core of this research programme, let us now consider its 'protective belt' of auxiliary hypotheses. Afterwards, I shall consider its positive heuristic.

---

[1] It may be argued that in disequilibrium situations the profit maximisation assumption is necessary to animate equilibrating adjustment but this is not so. The assumption of 'satisficing' or even 'bankruptcy avoidance' can serve the same purpose. What cannot be denied is that *some* motivational assumption is necessary. Cf. *below*, section 2 (*b*).

[2] Cf. *above*, p. 209.

Adherents of the neoclassical programme usually seek protection against apparent empirical discrepancies by claiming that the discrepant situation did not fully satisfy the conditions of applicability of the theory. For instance, it is pointed out that, contrary to what we supposed, adjustments are bound to be slow and that it is only in the 'long run' that perfect competition comes into its own.

Other typical examples are the following passages:

> . . . the effects of changes in wage rates, interest rates, tax rates, and so forth, are, if there is *effective* (my italics) competition, essentially independent of the relations among the various levels in the managerial hierarchy of the firm.[1]
>
> Competition . . . is sometimes called 'heavy', 'vigorous', or 'effective'. The simplest meaning of these adjectival modifiers is this: . . . if (the firm) is kept under continuing pressure to do something about its sales and its profits positions . . . the firm will not be able to pursue any objectives other than the maximisation of profits—for the simple reason that anything less than the highest obtainable profits would be below the rate of return regarded as normal at the time. . . . There can be no doubt about the fact that competition is not effective in many industries and that many, very many firms are not exposed to vigorous competition.[2]

Machlup's adjectival qualifications tell us nothing about the limits of applicability of the perfectly competitive model; they only tell us that the perfectly competitive model is only applicable where it can be applied: a classical *ad hoc* stratagem.

In order to clarify further the type of auxiliary hypotheses advanced by situational determinists, let us consider those which they are barred from using. They are barred, for instance, from proposing that the decision making units generally have 'slack' in their production and selling methods, and that this slack varies with aspirations and achievement; or that the utility function of the decision-maker may change in response to certain changes in situational conditions though remain unaffected by others. (Under perfect competition slack would be excluded altogether by the perfect knowledge assumption; in cases where slack obviously exists the proponent of perfect competition would say, quite logically, that his model is inapplicable.)

The 'positive heuristic' of situational determinism is its puzzle-solving mechanism. This involves the analysis of equilibrium conditions as well as comparative statics. It enables one to answer two interrelated types of questions: (*1*) Why do commodities exchange at given prices? (*2*) What are the effects of changes in parameters (say demand) on the variables of our model once adjustment has taken place? This positive heuristic has the advantage of generating statements about equilibrium conditions, and about changes in variables in response to variations in parameters, in the absence of exact quantitative information. The comparative static mechanism was first rigorously spelt out by Samuelson.[1] Archibald has argued however that this method yields testable conclusions only if very restrictive conditions are fulfilled, and that therefore its value for the empirical assessment of theories is very small.[2] Moreover, the comparative static apparatus cannot answer certain questions concerning the possibility of attainment of equilibrium or concerning the path followed during disequilibrium.[3]

---

[1] Machlup [1967], p. 15.

[2] *Ibid.*, p. 18.

(*b*) *Monopolistic Competition.*

Chamberlin's celebrated book *The Theory of Monopolistic Competition* illustrates very well the attachment of neoclassical economists to situational determinism. And so does its reception.[4] In a recent paper assessing the impact of Chamberlin's book, Samuelson writes: 'Marshall's crime is to pretend to handle imperfect competition with tools only applicable to perfect competition.'[1] But I suggest that it is more appropriate to say that, if a so-called crime was committed, it was Chamberlin's crime to pretend to handle imperfect competition with tools only applicable to perfect competition.

---

[1] Samuelson [1947], chapter II.

[2] Archibald [1961] and [1965].

[3] For a simple exposition see Lipsey [1971], pp. 135–6.

[4] The wide and immediate acceptance of Chamberlin's monopolistic competition theory may have been partly due to its apparent ability to resolve the protracted 'cost controversy' about the compatibility of decreasing costs and competitive conditions. This controversy was set off by J. H. Clapham's article on 'Empty Economic Boxes' ([1922]) and extended at least to the symposium on 'Increasing Returns and the Representative Firm' in the *Economic Journal* in June 1930. The alleged high point of this controversy consists of Sraffa's [1926] which is a skilful recantation of Marshall's (and indeed Cournot's) doubts concerning the monopoly-perfect competition package. Two major anomalies in the Marshallian programme were highlighted: (*a*) the inconsistency of perfect competition with falling costs; (*b*) the glaring unrealism of a perfectly elastic demand curve. Chamberlin's [1933] classic appeared to solve both problems while, at the same time, not demanding the wholesale abandonment of Marshallian orthodoxy. An excellent short review of the controversy is given in R. Robinson's [1971], pp. 17–23.

Although it contributed little to our understanding of entrepreneurial behaviour under competitive conditions, the theory of monopolistic competition is sometimes alleged to have produced incidental benefits. One is the recognition of the concept of marginal revenue. Another is the emphasis on the distinction between one firm and the collection of firms called an industry.

On the concept of marginal revenue there is little to be said. Admittedly it furnishes a unified way of describing the maximising conditions in neoclassical models. However, these conditions may be described in all cases without employing the marginal revenue concept at all. The concept has been neither heuristically nor empirically fruitful.

For the second point, some credit must go to the theories of monopolistic and imperfect competition. Before the publication of Chamberlin's and Robinson's books the distinction between the single firm and the industry was obscure. At times it was even suggested that the distinction was insubstantial, and that a device such as Marshall's representative firm could be considered as denoting a small scale replica of an industry. The literature of monopolistic competition has removed this particular confusion by insisting on the distinction between a firm and a group or collectivity of firms (an industry) and between the demand for the output of a firm and that of a collectivity of firms.

Even a casual inspection of what Chamberlin terms his 'large group' case reveals the obvious kinship with perfect competition. In both there is 'atomistic competition': the number of firms is so large that no action of any single firm can affect the price of the product and that no coalition is possible. In both, the producers face identical cost and demand curves. In both, entry is free and prompt and continues as long as profits are above minimum. In both, firms maximise profits by producing at the point where marginal cost and marginal revenue are equalised. The firms under monopolistic competition produce goods which are different in the eyes of the consumers but which do not demand any special knowledge or advantage (if the model is to be consistent) on the part of the producer who is responsible for their differentiation. Thus the only difference between perfect and monopolistic competition consists in slightly modifying the situational description; the hard core and the analytical machinery of the perfectly competitive paradigm are preserved intact. In equilibrium, in both cases, the individual firm's demand curve is tangential to its average cost curve. This means that only this particular price-output combination will keep the firm in business. If the demand curve cuts the average cost curve this means that profits will be made which will attract entry, thereby shifting the individual demand curve backwards until it is again tangential to the average cost curve (at which point there will again be no incentive to entry). In both cases, the situational description is such that optimising behaviour (yielding merely subsistence profits in equilibrium) is the only way of avoiding elimination from the industry.

Thus perfect and monopolistic competition share the common neoclassical 'hard core'; and monopolistic competition results from a slight modification of the situational assumptions of perfect competition.[2]

(*c*) *Duopoly and Oligopoly: Two Apparent Exceptions.*

Traditionally, the problem of oligopoly has been to analyse and explain economic behaviour in a market situation where there are few sellers and many buyers. The decisive characteristic of oligopolistic models is not numerical fewness as such, but 'fewness' in the idiosyncratic sense that decisions are interdependent.[3] In oligopolistic market situations the awkwardness of a single-exit model is most apparent. In such situations, suboptimal behaviour is still viable and the guesses and counterguesses about competitors' decision processes are crucial factors in the determination of price-output behaviour: we are confronted with genuine multiple-exit decision-situations. This being the case, one might assume that further psychological and social psychological considerations would have to be introduced to supplement the model. However, when we look at classical and neoclassical oligopoly theory, a very different picture emerges. Psychological or behavioural considerations often do not come in at all and when they do, as in the Cournot-Bertrand model, they are dragged in in order to mould the oligopolistic situation into a shape amenable to a single-exit solution: the situational or cognitive assumptions are tampered with in such a way as to reduce the oligopoly problem to a kind of monopoly problem with a single-exit solution.

---

[1] Samuelson [1967], p. 112.

[2] For an elucidation of the concept of 'situational assumption' see *below*, section 2 (*b*).

[3] In this idiosyncratic conception of 'fewness' it is important that industries with quite a large number of firms may nevertheless qualify for treatment by an oligopolistic model

Traditionally, the question *why* there are 'few' sellers has not been closely investigated by oligopoly theory. The theory simply assumes that entry is barred by institutional or physical means (for instance, by economies of scale or by privileged access to primary factors of production). The type of questions raised by oligopoly theory are typically of the traditional classical and neoclassical variety, notably questions relating to the determinateness of equilibrium and of the effects on the variables of changes in parameters. Indeed, the Cournot-Bertrand type of oligopoly[1] (they called it duopoly) reflects again the principal heuristic characteristics of the general neoclassical programme.

Let us first look at a *difference*. In oligopoly each decision maker is involved in outguessing the other and, with perfect knowledge on both sides, we would be led to an infinite regress. Thus, even when the assumption of perfect knowledge is retained concerning all other aspects of the market situation, oligopoly theorists have to admit one qualification, namely, incomplete knowledge about each others state of mind. But if perfect knowledge is impossible does not situational determinism break down? Is not the oligopolistic situation a multi-exit one? Traditional oligopoly theory has tried to avoid this unwanted conclusion. One way has been to restrict each seller's knowledge to the assumption that the other will go on producing the same amount whatever the first does. Now a single-exit solution is forthcoming. For now the duopolist may subtract his rival's output from the output indicated by the market demand curve at each price and thereby obtain his own demand curve. He will then (by

---

and that in general, the analysis of oligopolistic phenomena has wide applicability in the domain of market behaviour.

[1] Cournot [1838] and Bertrand [1883].

the usual marginalist calculation) maximise as a monopolist. But since the other duopolist, by assumption, proceeds on the same cognitive assumptions (and he too maximises), his output will change so that the first is maximising on (knowingly) false assumptions. A process of trial-and-error adjustments *may* lead in due course to a situation where each duopolist's assumption about the other's output is correct and neither will have an incentive to change output.[1] (Of course, the process of trial-and-error might not be convergent, and then no equilibrium would result). It is clear from this brief account that Cournot-type duopoly, though departing a little from the hard core of situational determinism, does so only to the minimal extent required to enable a unique price output decision which may be calculated on the basis of objective market data. The cognitive assumptions are modified only in order to restore the 'single-exit' type solution. There is no empirical justification for the modification.

Chamberlin's contribution to the oligopoly problem is even more in the tradition of situational determinism. He attempts to solve the oligopoly problem while maintaining all the assumptions (*i*), (*ii*), (*iii*) and (*iv*). His 'solution' is that with symmetry, product homogeneity and the maximisation-motive, the participants in an oligopoly are inevitably led to collusion and hence to a joint monopoly price. Yet there is no inevitability about this outcome. For collusion only implies all-round maximisation if there is a 'fair' division of the profits. But no convincing reason is given by Chamberlin or anybody else why we should expect this.[2] Chamberlin's solution adds nothing to the theoretical armoury of the oligopoly theorist, since it does not go beyond the analysis of pure monopoly. In this application of situational determinism, the big difference between a single decision maker and a coalition of two or more is concealed by an appropriately constructed situation. The interesting problems of duopoly (namely, why and on what terms do duopolists collude and how does this affect their decisions?) are accommodated *ad hoc* by being transferred into the premises.[3] Moreover, the structure of the Chamberlinian situation leaves the question of the division of the joint profit undetermined and presumably irrelevant to the stability of the equilibrium.[4]

---

[1] The Bertrand-Edgeworth solution of the duopoly problem is only trivially different from Cournot's: according to Bertrand, the cognitive assumption should refer to the constancy of price, not to that of quantity. This, Bertrand argued, led to a 'price undercutting' process and eventually to a 'zero profit' equilibrium.

[2] More recently Fellner ([1949]) has dealt with the problem of collusion mainly on the analytical level.

[3] Chamberlin [1933], pp. 100–4 of the 6th edition ([1966]).

[4] In his later work, collected in his 'General Theory of Value' ([1957]), Chamberlin tried to minimise the central role of the 'large group' monopolistic competition case and shifted the emphasis on his 'small group' situation. Yet his approach still does not come to grips with the oligopoly problem since he still cannot free himself of the blinkers of situational determinism.

Chamberlin's writings in price theory from 1933 onwards reflect the belief that a satisfactory theory of the firm can be developed simply by inserting elements of monopoly into a theory of market behaviour according to which firms operate under situational determinism.

It has come to be recognised in subsequent work (some interesting variants of which is noted in the concluding part of the present section) that even when sellers are relatively numerous decisions are often interdependent. The inadequacy of situational determinism is especially manifest when decisions are interdependent. Recognition of the importance of interdependence in decision making has therefore probably contributed to the partial retreat from situational determinism in the recent literature on the theory of the firm. Unfortunately, the rising interest in the problem of oligopoly has not always been accompanied by a preoccupation with the problem of interdependent decision-making. This is most clearly seen by a brief excursion into a recent oligopoly theory proposed by J. S. Bain and (though in a slightly different version) P. Sylos-Labini.

Classical (Cournot) and neoclassical (Chamberlin) oligopoly theory 'explains' the small number of sellers in terms of exogeneous barriers to entry. Chamberlin himself was not particularly concerned with the conditions of entry and the emergence of oligopoly, being more concerned with price-output determination. More recently, however, theorists have focused on the conditions of entry as important situational conditions both in creating oligopolies and in determining price and output.[1]

The recognition of the importance of potential entry into an industry has not gone unnoticed by classical and neoclassical writers. Both Marshall and J. B. Clark, among others, realised that under certain circumstances the threat of potential entry may have the effect of keeping price below the monopoly level.[2] Yet these writers treated potential entry under monopolistic conditions as an anomaly and never attempted to incorporate it into neoclassical theory. Bain ([1949]) stressed the condition of entry into an oligopolistic industry and suggested that it is an important situational condition in generating a determinate price and output. Bain however does not come up with any specific model. He assumes a long-run equilibrium price at which entry will not occur, but does not tell us how it is to be derived.

The first specific model which incorporates Bain's idea comes with Sylos-Labini's [1956] with refinements by Modigliani ([1958]). In the present instance the 'Sylos' model is of particular interest since its highly simplified character allows a situationally determinate, single-exit solution.

In its simplest and most amenable form the Sylos model involves a homogeneous oligopoly model where all firms face the same cost and demand conditions, with average costs decreasing as output increases. The members of the homogeneous oligopoly are assumed to act collusively in the face of potential entrants to the industry who threaten to reduce profits by making their entry effective. Thus, it appears that the Sylos model shares the hard core of other neoclassical models in maintaining the implicit rationality assumptions (maximisation, perfect knowledge, perfectly functioning market) and modifying only the situational detail. Moreover, since decision inter-dependence (as opposed to collusion) among existing producers is again assumed away it becomes debatable whether it is a genuine theory of oligopoly at all. We may just as well consider it as an extension of Marshallian monopoly theory.

---

[1] Cf. Sylos-Labini [1956], Bain [1956]. For a more recent contribution see Dewey's [1969].

[2] Clark [1907], pp. 380–81 and Marshall [1919], chapter I, section 1. A clear statement of the importance of the conditions of entry for price and output determination is made in Kaldor's [1935], footnote 18.

Price and output are uniquely determined in the Sylos model in the following way. The existing producers behave as a single monopolist threatened by new entry. They increase their output above the pure monopoly level to a so-called 'limit output'. At this level, the price ('limit price') is such that: (*a*) *They* make a (reduced) monopoly profit; (*b*) any new entrant (who would have to start with higher average costs) would make a *loss*, given that the existing producers maintain their present output.

The *strict* solution is achieved at the expense of stringent situational assumptions in addition to the usual neoclassical assumptions. It requires for instance that: (*a*) there must be an unlimited number of potential entrants; (*b*) a potential entrant must be able to enter instantaneously; (*c*) barriers to entry must be either due to economies of scale, or to absolute cost advantage, enjoyed by established firms over potential entrants. Moreover, we require an assumption concerning the beliefs of the potential entrants about what the existing producers will do. This particular assumption, christened (by Modigliani) the 'Sylos Postulate', states that the potential entrant assumes that the existing producers would maintain output, reducing price to the extent necessary to enforce this output policy. Given this belief on the part of existing producers, the latter are, according to Sylos-Labini and Modigliani, led to a unique price-output combination determined by the relevant cost and demand conditions.

Several problems arise with the 'Sylos' model. It is not clear why the existing producers should communicate the threat of enforcing losses on the entrant by actually implementing the Sylos policy before entry takes place. Surely, ways could be found of communicating the threat of exploiting the entry barrier while simultaneously maintaining a monopoly price and output. Besides, it is not clear what motivates the existing producers. The trivial Sylos postulate does not provide us with the answer. If we retain the standard assumption of profit maximisation we may find that under a wide range of circumstances it conflicts with the Sylos pricing policy. Further specification of the decision makers' knowledge and motivation is required. It is characteristic of the tenacity and the pervasiveness of situational determinism that neither Sylos-Labini nor Modigliani recognises the opportunity open to them to investigate alternative assumptions concerning the motivation, cognition and decision procedures of inter-acting potential and actual competitors. According to Modigliani: 'it is both to Bain's and Sylos's credit that [they have moved] us away from conjectural variations and similar subjective notions and [focused] instead on objective market and technological data . . . .'[1] By 'subjective notions' he presumably means the interdependence of decisions among existing producers and the interaction—by means of threats and counter threats[1]—between existing producers and potential entrants that are excluded from the Sylos model.

---

[1] Modigliani [1958], p. 222. Modigliani adds that in doing this 'they have provided us with theories rich in empirical content and capable of being disproved by evidence' ([1958], p. 222). However this claim is disputable. The difficulties of subjecting the 'Sylos' model to actual tests are described by Osborne [1964]. Recent efforts to improve the theory have been made by Pashigian [1968].

2 STAGNATION OF THE NEOCLASSICAL PROGRAMME OF SITUATIONAL DETERMINISM AND EMERGENCE OF RIVALS

None of the various neoclassical theories of the firm is satisfactory. Not that they are, or should be, embarrassed, by various empirical anomalies. Indeed, such anomalies were already known to Adam Smith, to Cournot, and to Marshall,[2] and they were imperturbed by them. It is not their *empirical* satisfactoriness that I shall question in this section. The question I shall raise is whether, during the last few decades, the neoclassical programme for the theory of the firm has been theoretically progressive or degenerating. To answer this question we must pay special attention to the theory of monopolistic competition. It is widely held among economists (with the notable exception of the Chicago School) that this theory was an important advance within the neoclassical framework. I shall argue that it was actually a degeneration from Marshall.

(*a*) *The Appraisal of Perfect and Monopolistic Competition.*

The perfectly competitive model represents the most impressive application of situational determinism. This model, in its simplest, static form is logically impeccable; and has some explanatory, predictive power: for instance, exogeneously caused changes in costs can be expected to result in corresponding changes in prices and in influx or exit of firms from the industry (until equilibrium is restored). Yet the simplicity and ideal character of the model makes it difficult to locate and identify anomalies; for the fundamental assumptions and the initial conditions which have to be fulfilled, if the model is to be tested, are never exactly realised; and there is little guidance to indicate how much divergence is tolerable.[3]

For instance, it has been found empirically that firms in highly competitive industries sometimes fail to behave in the predicted way in the face of large changes in cost. There may be substantial changes in costs *without* substantial entry or exit of firms.[4] The programme can however

[1] It is unfortunate, though hardly surprising, in view of the dominance of situational determinism, that economists persist in ignoring psychological literature on bargaining behaviour under conditions of 'threat'. See, for instance, Deutsch and Krauss [1960] and Kelley [1965].

[2] Cf. *above*, p. 208, footnote 2 and p. 213, footnote 4.

[3] 'The reason for not stating the weakest assumptions . . . for competition is that they are difficult to formulate and in fact are not known precisely' (Machlup). For a brave attempt to state the limits of applicability of monopolistic competition theory see Machlup [1952], pp. 315–16.

[4] See, for instance Bauer [1945]. This example refers to the rubber industry in Malaya which appeared to fulfil the initial conditions for perfect competition unusually well.

attempt to accommodate such difficulties: the blame can be put on the auxiliary (situational) assumptions, and the hard core can thus be saved. But these adjustments have never led to anything but to *ad hoc* adjustments. For many this came as no surprise. After all, it could be easily seen that the perfectly competitive model, by its construction, could only deal with a narrow range of questions of a 'comparative static' type.

But one useful heuristic function is certainly left for this idealised model. If its mathematical techniques can be fruitfully extended to different market situations, then it may be claimed that it has accomplished the valuable task of being a *testbed for the development and application of mathematical methods* to the theory of the firm. This I take to mean a certain kind of theoretical progress.

Unfortunately, the extensions of situational determinism to monopolistic competition and oligopoly do not have even this merit. Chamberlin's introduction of differentiation is a rather radical step towards executing the heuristic of situational determinism. Instead of using the argument of product heterogeneity as an escape clause (to indicate that whenever heterogeneity is present, perfect competition is only imperfectly applicable), he incorporates product differentiation in his model itself.

The solution of several problems was expected from this modification, for instance, the problem of selling costs and that of the decreasing cost equilibrium which were both excluded by perfect competition. It was also expected that the modified theory would lead to a wealth of testable consequences. But Chamberlin himself was quite uninterested in the derivation of testable predictions, and later efforts by Kaldor, Demsetz and Archibald have shown that—on the qualitative comparative static level—the model is even poorer than that of perfect competition. The reason is common to monopolistic competition and to monopoly theory: if we vary a parameter (demand), the change of the variable we wish to predict (price) may be construed as either an increase or a decrease, depending on the relevant quantitative information. However, no such quantitative data are available and in their absence one cannot deduce even qualitative predictions.[1]

The only novelty offered by Chamberlin is, then, the excess capacity theorem: equilibrium must occur at a point where costs are decreasing. Yet serious doubts have been cast on the validity of this theorem. The first serious difficulty was highlighted by Kaldor who pointed out that Chamberlin's large numbers case with zero profits and free entry converges to perfect competition. For as newcomers enter the industry they simultaneously reduce profits and render the individual demand curves more elastic. Admittedly, equilibrium may occur while profits are still being made if the situation is such that the entry of one more firm would convert minute profits into losses. (Even if we accept Chamberlin's 'heroic assumptions' of symmetry and uniformity.) Yet Chamberlin discards small profits ensuing from differentiation, calling them zero, while continuing to stress the slight divergence of cost from the minimum point of the average cost curve as a major result of the theory. These qualifications are, however, of the same order of magnitude and we cannot disregard the one without also dismissing the other. But if both are dismissed we are back to perfect competition. If, on the other hand, we admit small monopoly profits (as seems reasonable) then the equilibrium is analytically identical with that of pure monopoly. In both cases the novelty of the tangency equilibrium is lost.[1] Another serious criticism of the decreasing cost equilibrium came from Demsetz. He claimed that if we include in the firm's cost function any demand-increasing costs whatsoever and allow them to vary optimally with output—as indeed we do in the case of production costs—then decreasing cost equilibrium is no longer assured: it may occur and it may not.[2]

---

[1] See Kaldor [1935] and [1938], Archibald [1961], Demsetz [1964].

The only remaining way of testing the theory is to attempt to observe the equilibrium conditions. Yet such attempts are beset with difficulties. If, for instance, we want to test zero profits as an equilibrium prediction we have to decide how small profits must be to be considered 'zero'. Then, of course, we are still left with the problem of the reasonableness of our convention.

Finally, there are serious logical difficulties with Chamberlin's notion of a product differentiating group. These difficulties were first intimated by Triffin[3] who pointed out that the assumption of free entry was inconsistent with that of product differentiation, and that there was no criterion to delineate a product differentiating group or 'industry' from the multitude of firms competing in an economy.

It appears therefore that the Chamberlinian 'large numbers' case fails to fulfil the testability criterion. It is, therefore, less criticisable even than perfect competition. It even fails, as we have just seen, the requirement of logical coherence.

In the light of all this it is amusing to look at a few quotations from 'specialists' in the field: J. S. Bain writes of Chamberlin's theory: '. . . in

---

[1] For a critical discussion of the tangency equilibrium see also Machlup [1952], chapter 10. When the 'heroic assumptions' are relaxed it no longer follows that there should be a decreasing cost equilibrium with zero profits. See also, Chamberlin [1966], p. 195. Under monopoly we may have equilibrium with decreasing costs and positive profits but whether or not this is the case depends on the position of the cost and demand curves.

[2] Cf. Demsetz [1964].

[3] Cf. Triffin [1940], pp. 88, 118.

"Monopolistic Competition" Chamberlin really introduced a new price theory with a vastly greater empirical relevance than that of preexisting theory and with an immensely increased immediate or latent power to generate hypotheses concerning enterprise behaviour'.[1] P. A. Samuelson calls monopolistic competition: 'the best current model of price theory'[2] and Robert Bishop writes: 'As I judge the consensus of economists, Chamberlin's "Theory of Monopolistic Competition" and Mrs. Robinson's "Economics of Imperfect Competition" are acknowledged to have touched off, in 1933, a theoretical revolution whose relative importance in the micro-economic area was comparable to that of the Keynesian analysis in macro-economics.'[3]

I have suggested that the dominant idea in the theory of the firm is 'situational determinism' and that some of the major debates, especially the Chamberlin-Chicago controversy, are mere family quarrels between slightly different variants within the same programme. I have also suggested that this programme after its initial progress in the field of perfect competition, was overextended into fields in which it failed. Before I come to the competing programmes that are beginning to emerge, I will offer a more detailed discussion of one postulate in the hard core of situational determinism: the Rationality Principle.

(*b*) *Popper's Rationality Principle.*

I have argued at length that the classical and neoclassical programme for explaining business behaviour has been characterised by the fundamental regulative assumption that decision making agents or agencies optimise under severe objective constraints, and therefore have virtually no *Spielraum.*

---

[1] Bain [1967], p. 149.

[2] Samuelson [1967], p. 112.

[3] Bishop [1964], p. 33. These appraisals are invalid and the sociologist of knowledge will be interested why. Irrational appraisals can be explained by 'external' causes. (For this 'internal-external' distinction cf. Lakatos [1971*a*] and Kuhn [1971].) One such 'external' explanation could be that if capitalism contains a lot of monopolistic competition then it would be shown to involve social waste through excess capacity.

Excess capacity under monopolistic competition was at time instanced as a defect of the market system as it actually works. Chamberlin thought that excess capacity was not necessarily wasteful and that within his model it represents the price that consumers were willing to pay for product differentiation. Unfortunately, this explanation does not carry conviction since, in Chamberlin's model, differentiation is more or less imposed on the consumers. Nicols ([1947]) has criticised Chamberlin for requiring a rather perverse consumer theory in order to carry through the analytics of his 'large group' case.

We may note in passing that unused capacity is often not genuinely excessive or is excessive only on certain arbitrary criteria. Unused resources often yield the valuable service of availability as is evident from people's readiness to own books or clothes instead of relying on lending libraries or dress-hire firms. Similarly, shop assistants, who are not constantly busy, are nevertheless not redundant as otherwise they would be dismissed.

In order to show the inexorable logic of a competitive situation I shall spell out in some detail the explanatory schema of a producer's decision to produce a certain output (say Q_x) under 'normal' equilibrium conditions:

EXPLANANS	Psychological Assumption	(*1*) Producer y (of commodity x) strives to maximise profits
	Situational Assumptions	(*2*) Producer y is operating under perfectly competitive conditions
		(*3*) The price of x is P_x
		(*4*) y's marginal cost curve is MC (the curve showing the addition to total cost, for any output, of an additional unit of output).
	Cognitive Assumption	(*5*) Producer y has perfect knowledge of his situation as characterised by (*2*), (*3*), (*4*).[1]
RATIONALITY PRINCIPLE		(*6*) y acts appropriately to this situation
PREDICTION (EXPLANANDUM)		(*7*) y produces output Q_x such that $MC = P_x$[2]

It is part and parcel of this approach that the constraining part of the decision making situation is that pertaining to the situational assumptions. Indeed, it could even be said that in the simple, static case of the perfectly competitive situation the psychological assumption, namely, that of profit maximisation, becomes almost superfluous.

I have shown earlier[3] that if a firm (under perfectly competitive conditions) acts 'appropriately' it can act only in *one* predetermined way. Thus the motivational assumption (*1*) may be weakened from profit maximisation to bankruptcy avoidance, without affecting the derivation of the prediction.[4]

In this neoclassical explanation-schema, the decision maker's psychological characteristics become almost irrelevant. It suggests economic theorising without psychology. Indeed, economics becomes essentially *autonomous*. This was first recognised and argued for by Marx. Marx laid great stress on the impersonality of economic forces and on the strict constraints of the objective economic situations which work in history independent of individual awareness or individual action. Although Popper attacks Marx's holism he endorses (with acknowledgements) the idea that psychological laws are irrelevant or trivial for economic analysis[1]:

---

[1] Popper includes (*5*) in the situational assumptions, cf. *below*, p. 224, footnote 4.

[2] The schema may be put in entirely subjective terms. In this version (*2*), (*3*) and (*4*) are replaced by the producers corresponding beliefs and (*5*) becomes superfluous. However, subscribers to methodological individualism may be inclined to keep certain objective statements in the schema; firstly because this may render it more testable and secondly because social events which come about as unintended consequences can often not be explained in purely subjective terms.

[3] Cf. *above*, p. 211.

[4] In the case of pure monopoly, the assumption of bankruptcy avoidance will not yield a unique price output solution. Here profit maximisation *is* required.

> We may say that our actions are to a very large extent explicable in terms of the situation in which they occur. Of course, they are never fully explicable in terms of the situation alone; an explanation of the way in which a man, when crossing the street, dodges the cars which move on it may go beyond the situation and may refer to his motives, to an 'instinct' of self-preservation, or his wish to avoid pain *etc.* But this 'psychological' part of the explanation is very often trivial, as compared with the detailed determination of his action by what we may call the 'logic of the situation'; and besides, it is impossible to include all psychological factors in the description of the situation. The analysis of the situation, the situational logic, plays a very important part in social life as well as in the social sciences. It is, in fact, *the* method of economic analysis.[2]

But entrepreneurial behaviour is usually very different from street crossing. For crossing the street we need little imagination, skill and deliberation. For entrepreneurial activity one often needs a high degree of imagination and skill; and these are processes which can hardly be reflected by crude single-exit models.

A central feature of Popper's situationally deterministic, antipsychologistic programme is his so called 'rationality principle'. It is difficult to extricate from its main proponents (Popper and Watkins) its precise logical and empirical status and function in the explanatory model. In his [1967] discussion Popper intended to dispell the 'countless misunderstandings' of this principle.[3] From this recent paper, it turns out that Popper's position differs much more from Marx's than one would have thought from reading *The Open Society*. For in this paper, Popper adds: 'We must remember, of course, that the situation as I use this term already contains all the relevant aims and all the *available relevant knowledge*, especially that of possible means for realising these aims.'[4] But he still does *not* include into the 'situation' irrational beliefs: 'There are obviously vast personal differences not only in knowledge and skill—*these* are part of the situation—*but* in assessing or understanding a situation; and this means that some people will act appropriately and others not.'[5] But why exclude from the 'situation' irrational beliefs and why not separate the problem of irrational belief from the problem of irrational action? Watkins subjectivises the 'situation' to include all beliefs and consequently weakens Popper's rationality principle and christens it 'imperfect rationality principle'.[1] But why not then go one step further and include 'unconscious motives' into the 'situation'?[2] One answer could be because this would lead to psychologism.

---

[1] Popper [1945], chapter 14, [1957], pp. 152–8 and [1967], p. 144.

[2] Popper [1945], vol. ii, p. 97; my italics.

[3] His effort was not quite successful: he describes the principle as non psychologistic, not *a priori* true, as irrefutable, as 'almost empty' and yet not tautologous; and finally it is described 'trivially true' and as 'clearly false'. Popper [1967], p. 144–9.

[4] Popper [1967], p. 144; my italics.

[5] *Ibid.*, p. 145.

Let me now attempt to spell out in some detail the role of the rationality principle[3] in this form of explanation. This may be carried out in two steps: (*1*) I shall first consider a naive view of this principle as a psychological/neurophysiological bridge—principle—an interpretation which has unfortunately caused some unnecessary controversy.[4] (*2*) I shall then consider the rationality principle as a 'decision principle', first under conditions of perfect knowledge, and then under imperfect knowledge. I shall be arguing that under imperfect knowledge the rationality principle really amounts to a descriptive psychological or social psychological decision theory.

(*1*) What I call the 'bridge' role of the rationality principle is the assumption that the decision maker behaves adequately or appropriately to his 'decision schema'. (This is the conjunction of the agent's beliefs and his preference map.) According to Watkins, once the agent has calculated the response that accords with his preferences and his situational information, then the rationality principle tells us that he actually *does it*. Under this interpretation, the rationality principle cannot be subsumed under either the evaluational or the situational assumptions since it remains invariant as the situational and evaluational assumptions change. Such a rationality bridge-principle, to my mind, ought to have some psychological and physiological content in that it asserts a stable connection between a certain state of mind (intent to implement a decision) and appropriate motor responses to realise this intent.

The defenders of the 'bridge principle' view agree that there is a need for a bridging premise between decision schemas and action but they refuse to admit its psychological character. But then what content does it have? Watkins is at a loss. In his [1970] he explored three charges usually levelled against the rationality principle: the charge of *triviality* the charge of *falsity* and the charge of *redundancy*.

---

[1] Watkins [1970].

[2] Cf. Hempel [1965], pp. 483 ff.; see also Watkins [1970], pp. 174–5, 208.

[3] There is a bulky literature on the philosophy of action which deals with the problem of rational behaviour and with general principles of rationality as they feature in the explanation of human action. See, for instance, Ryle [1949], Anscombe [1957], Davidson [1963] and Hempel [1965]. For my part I shall be concentrating on the role of the Popperian rationality principle in deductive explanations; and this to the extent necessary for clarifying the situationally deterministic programme in economics.

[4] See, for instance, Donagan [1970].

Consider the charge of *triviality*. The 'bridge principle' version of the rationality principle simply amounts to the assertion that intentions to act in a certain way are (circumstances permitting) translated into appropriate actions. One wonders whether this is what Popper meant when he called the principle 'almost empty'[1] or what Watkins meant when he wrote that the rationality principle has the 'sort of triviality that goes with indispensability'.[2] If so, the rationality principle would be a fairly trivial principle from the point of view of the social scientist who, in order to proceed with his investigation, assumes a stable connection between a state of mind (intention to respond) and physiological response.

Let me now turn to the charge of *falsity*. Watkins defends the principle by furnishing examples of apparently recalcitrant instances that *do not* constitute criticisms of it. For on Watkins's weakened version of the principle, 'mad', 'irrational' or 'stupid' action may nevertheless turn out to be explainable as (imperfectly) rational in terms of some ingeniously reconstructed decision scheme plus the imperfect rationality principle, and so do not constitute violations of it. Yet the conditions (if any) which *would* violate even this imperfect rationality principle remain obscure. In my 'naive' interpretation the principle *may be false* but it can be reasonably included in the ceteris paribus clause and is of little interest—I am tempted to say irrelevant—to the social scientist.

Finally, let us consider the charge of *redundancy*. The conclusion we reach depends on how we interpret the premises in the explanatory schema. If the propositions expressing the so-called aims in the explanation schema are represented by a behaviourist preference map and the situational information by appropriate stimulus circumstances then the resultant practical conclusion also implies (deterministically or probabilistically) a certain definite action. In this case the rationality principle *is* redundant.

If, on the other hand, the 'aims' are viewed as motivational assumptions interpreted in accordance to modern physiological psychology then motivation and the physiological process by which it is translated into overt response form a single closely interlinked system. Indeed, Hebb ([1955]) suggests that the motivation of an organism implies a physiological state of 'arousal' where the appropriate parts of the brain are ready to commit the body to some type of response without the precise physical character of the response being specified. Indeed, there may be a considerable time-lag between the physiological state of arousal and the actual 'firing' of the nerve cells and, more importantly, arousal may occur without being followed by overt response. Thus this conception of motivation implies not only preference but also the initiation of physiological response

[1] Popper [1967], p. 144.

[2] Watkins [1970], p. 173.

in accordance with it. Profit maximisation, and the like, far from being primary psychological drives or motives, are rather institutionalised motives which originate from complex interactions between the social, physical and psychological environment of decision makers. And we do not have to spell out how these higher motives have come about in order to provide successful explanations. Still, according to the standard meaning of motivation in psychology, motivational assumptions are considered sufficient to animate action in the model, and then the bridge (rationality) principle is superfluous. It is however possible to give such interpretations to the concepts of motive, preference and situational appraisal that they do not by themselves imply appropriate action. Then we would need some psychological/physiological bridging principle. Thus whether or not the principle is redundant depends on our interpretation of the motivational and evaluational premises.

(*2*) The psychophysiological principle just discussed above bridges the gap between decision and behaviour. But a further gap may appear in an agent's decision schema: the one between situational appraisal and decision. In this case the cognitive status of the decision making agent is of primary importance since situational information, appraisal and selection depend on the kind of knowledge the agent has with respect to his preferences and his environment. This further gap may be seen to necessitate a second rationality principle.[1]

Let me first attempt to elucidate this second rationality principle under conditions of perfect knowledge.

Under the standard neoclassical assumption of perfect knowledge the process of reaching a decision and the process of carrying it out are fused. The reason for this is simple: once a perfect decision maker with a perfect utility map in a perfect market is given perfect information about his situation, then the decision is as good as made. The ranking of alternative outcomes is known and the decision associated with the best outcome is automatically selected. And by taking utility or preference maps as given, neoclassical economists have evaded the problem of the formation and changes in the preference maps.

There is one rather trivial qualification to the above analysis. It may happen, even under the assumption of perfect knowledge, that different decisions would lead to equally preferred outcomes. The rule 'select the most preferred outcome' does not in this case uniquely determine the decision. Thus if the decision maker wants to select one among several decisions leading to equally preferred outcomes he must adopt a selection procedure which goes beyond the simple optimising rule. He may impose any other idiosyncratic rule (random devices) which may well be adhered to in all cases where the optimising rule does not lead to a unique decision. Such rules may be generalised in the following way: 'In situations of type x decision makers of type y employ process z to select among alternative optimal outcomes.'

---

[1] I owe this distinction to John Watkins.

In general, in neoclassical microeconomics, the gap between the agent's situational appraisal and the decision he makes is bridged by the optimisation principle which, as I have argued above, is trivialised under perfectly competitive conditions. If viewed as an optimising (or some equivalent) principle the rationality principle is a logically indispensable premiss (though it could be considerably weakened under certain very special conditions); and, in general, far from being analytically true it is empirical, notably with some psychological/physiological content.

Decisions are of course generally taken in conditions of less than perfect knowledge and less than articulate or even consistent preference maps. But then the decision the agent will reach depends not only on his preferences and situational picture but also on his *decision process*, namely, the way in which he will process the incomplete information available to him.

In general there will be an indefinite variety of alternative decision processes for the same situational and motivational picture and each may lead to very different decisions being adopted. The actual decision may depend on more or less idiosyncratic rules or on rather stable rules related to various personality traits of the decision makers or to their social background or to other social-psychological forces.

Under imperfect knowledge (risk, uncertainty, ignorance) there is an important gap between motivation, preference, situational assumptions and reaching a decision. In explaining decision behaviour we may formulate generalisations about such *procedural rules* (canons for dealing with risky situations, heuristic procedures, *etc.*) for decision selection. These generalisations will then constitute central factors in the decision making process.[1]

Let us recapitulate: in a decision model with perfect knowledge the second rationality principle relating to decision formation and selection is rendered otiose. In a model with the usual neoclassical assumptions (maximisation, perfect knowledge, perfect market) there is simply no room left for irrationality; merely for non-conformity to the assumptions characterising the model. Here the so-called rationality principle is absorbed into the postulates characterising the *homo economicus*. But such models have little application to actual entrepreneurial decision making.

---

[1] Hempel seems to consider a general descriptive psychological principle as an indispensable component of any explanation of behaviour but it is not clear what its empirical content or methodological status would be. Cf. Hempel [1965], pp. 468–9. For a discussion of the decision making process and in particular of decision procedures such as *search processes* and *heuristic processes* see Simon [1960] and [1972].

(*c*) *The Emergence of Economic Behaviouralism.*

The adoption of antipsychologism as a heuristic canon is not only unnecessary but, by restricting permissible types of explanatory generalisation, may halt progress in microeconomics. Indeed, as we shall see, an interesting rival to the neoclassical programme is now calling for modifications of the prevailing neoclassical theory of rationality. This may become clearer if we briefly consider certain new directions in which the theory of the firm has been moving, notably recent theories associated with Simon, Baumol and Quandt and Cyert and March. Simon writes: 'Oligopoly theory today is embarrassed by its inability to choose among the rich assortment of alternative assumptions that are available to it. It is unable to choose because the alternatives represent conflicting *empirical* statements about the *psychological* characteristics of economic man, and economic theory is not accustomed to deriving his characteristics from empirical observations.'[1] Simon mistakes situational determinism for apriorism and *psychologically* observation-motivated theories for inductive ascent. (For Simon's inductivism cp. especially his [1968].) Still, such (or even worse) philosophical errors do not reduce the challenge to situational determinism.

According to this emergent programme, instead of attempting to explain the behaviour of economic agents as best decisions in a constraining situation we should attempt to explain them as more or less good (or possibly disastrously bad) solutions in fluid and partially known or even completely misunderstood situations. It is clear that in such an approach the burden of explanation shifts to the behavioural characteristics of the acting agent. The psychology of learning, of attention and of motivation, and even lower level behavioural generalisations become relevant once the hard core of the neoclassical programme is put aside.[2] It is perhaps not premature to give a name to the cluster of theories that have emerged from this approach: my own choice is 'behaviouralism'.[3] This term is my own addition to methodological jargon. I trust it will not be confused with 'behaviourism' in psychology.

---

[1] Simon [1957], part IV, p. 198, my italics.

[2] For a sample of the relevant psychological literature see: James [1891], chapter XI, Hebb [1949], Thibaut and Kelley [1959], Kelley [1965], Simon [1967], Mackworth [1970].

[3] One may discern a third kind of approach to the theory of the firm: 'organisationalism'. In so far as it constitutes a separate programme the organisational approach lays the stress on the irreducible and constraining nature of institutional structures, properties and needs. In this light individual aims and characteristics could be seen as largely neutralised by such general (superhuman) organisational tendencies as 'integration' and

A simple illustration of a breakdown of the neoclassical programme, where the new approach may be helpful, is the case of bilateral monopoly. This is a market situation consisting of one buyer and one seller. If endowed with the usual neoclassical assumptions, the analysis of this market situation reveals that no determinate price-output behaviour can be predicted. Instead, we have a range of possible price output combinations.[1] Having presented the technical aspects of this result, Baumol concludes in his textbook:

> Beyond this it is difficult to narrow down any further the possible location of the final equilibrium point. Several possible suggestions have been offered—for example, the joint maximum point.... However, *it is difficult to see why* one may expect that *the bargainers should always be expected to end up at any one such point*. For that reason, many economists have concluded that the bilateral monopoly problem is indeterminate.[2]

As we can see from the italicised passage of the above quotation there appears (according to Baumol) to be no particular reason why the bargainers should choose any one of the infinity of alternative positions. In other

---

'adaptation'. Yet institutions not only have a tenacious structure (of internal and external relationships) which shapes individual behaviour but also appear to have needs and aims which are not reducible to those of individuals. Selznick writes: 'There are certain needs generated by the organisation itself, which command the attention and energies of leading participants' ([1949], p. 10). It can therefore be seen that the organisational approach is, in important ways, distinct both from the behavioural and the situationally determinate approaches. It perhaps may best be seen as a form of structural functionalism. Some important recent proponents who have stressed its relevance for economic theory are Simon [1945], Thirlby [1952], March and Simon [1958], and Leibenstein [1960]. Leibenstein, in his [1960] writes: 'Decisions are not made by individuals with entirely undefined parts to play in an organisation. In other words, it seems quite natural to view decisions as being made within the roles played by individuals who are members of an organisation' (ch. 9, p. 120). Later he writes: 'The firm as an organisation is to be looked upon as a collection of related roles' (ch. 12, p. 168). On this view of the internal structure of the firm, the structure and meaning of roles antedates the role-fillers. This characterisation of the organisational approach does not of course preclude the blending of some of its elements with behaviouralism—a task undertaken by Cyert and March ([1963]) and Williamson ([1970]). Yet at the present stage of development it might have been wiser to keep them apart. The behavioural approach may after all be applied to the single, isolated decision maker of neoclassical economic theory—there need not be a complex intrafirm structure for behaviouralism to be applicable. Indeed, it may be advisable to develop behaviouralist ideas, first in their application to a single decision maker and later to a complex decision making structure. The combination of behaviouralism and organisationalism in Cyert and March's textbook has not produced many results.

The organisational approach though shedding some light on the firm's internal structure and its effects on decision making has done little to solve the main problem of the neoclassical theory of the firm namely that of decision behaviour under conditions of competitive interdependence.

[1] For a short discussion see Baumol [1965], pp. 325–7.

[2] Baumol, *ibid.*, p. 327.

words neoclassical situational determinism does not offer reasons why one position should be the equilibrium rather than another. This result, far from uncovering an intrinsic indeterminacy in all market situations which fit the bilateral monopoly form, has rather 'uncovered' a weakness of the neoclassical approach. Stigler admitted this when he wrote: 'Indeterminacy has a special meaning: the conditions of cost and demand are not sufficient to determine the price and quantity. . . . To say that a situation is indeterminate is a refined way of saying that it is not fully understood.'[1]

Thus the problem of explaining some *actual*, *specific* price-output combination in bilateral monopoly situations is insoluble only within the neoclassical programme. If we are prepared to introduce psychological hypotheses concerning bargaining situations plausible solutions are readily available.

To take a recent example: consider the following (admittedly low-level) behavioural hypothesis which was tested under laboratory conditions by Johnson and Cohen[2]: 'Business oriented actors in situations of bilateral monopoly tend to resort to an equal split joint pay off.'[3] The results of the experiments carried out by Johnson and Cohen suggested that bargainers of this particular orientation and social background chose the optimal quantity at the intersection of their marginal schedules and negotiated prices at an equal split level.

The value of such experiments is for the present dubious and the status of the hypothesis, even if it proves independently testable, is precarious unless it can be integrated into a comprehensive research programme.[4] Still, it should be noted that such low-level behavioural rules have performed quite well in prediction and interesting results are reported by Baumol and Stewart[5] (in prediction of retail prices).

With hindsight, the classical Cournot-Bertrand, Stackelberg solutions for duopoly could be fitted into the behavioural camp. For they proceed by modifying the cognitive assumptions of classical theory, thereby tampering with the hard core of situational determinism. But Cournot and Bertrand use these modifications as a means of reducing the problem to one with a single exit, profit maximising solution. That is why I previously included them in the situational determinist camp. (Stackelberg is here the exception: he considers his indeterminate result as uncovering an intrinsic indeterminacy in oligopolistic market situations, *i.e.* 'price wars').

A behaviouralist theory of rationality, drawing both on the psychology of Kurt Lewin[1] and on ideas originating in engineering and cybernetics, has been championed by H. A. Simon.[2] This theory borrows from cybernetics the idea of a closed-loop feedback control system. The mechanism has a criterion θ_m by which it judges the situation. The mechanism has the ability to register the actual situation, θ_α, and thereby the difference $(\theta_m - \theta_\alpha)$. The control mechanism may then either amplify the difference, diminish it or leave it unchanged. An adaptive mechanism, which seeks to reduce this difference, is assumed by Simon. Following Lewin, a fixed criterion of achievement may be replaced by a flexible aspiration level[3] according to which the criterion θ_m is a variable depending on the achievement of the control aspect of the mechanism. If the corrective action of the control mechanism is successful in terms of approaching θ_m then θ_m rises. If the reverse is true it falls.[4]

---

[1] Stigler [1966], p. 207.

[2] Johnson and Cohen [1967].

[3] The phrasing is mine.

[4] Lakatos calls such propositions *ad hoc*$_3$. (Cf. his [1970], p. 175.) Kuhn might find them characteristic of the pre-paradigmatic period following a crisis.

[5] Baumol and Stewart [1971].

The non-optimising monopoly model is[5] a promising case for the application of Simon's model of 'satisficing' rational choice. We assume a single decision maker (monopolist) with no well specified objective function who is, however, able to rank potential outcomes as 'satisfactory' or 'unsatisfactory'. The model is a simple dynamic one consisting of a sequence of one-period decisions in which the situation at the start of each period is affected by the firm's past behaviour. Moreover, the motivation to make satisfactory profits varies directly with the actual results achieved. By the appropriate algebraic techniques and by the introduction of random variables, a general non-optimising monopoly model can be constructed. This may then provide a series of price-output decisions by the utilisation of computer simulation techniques. Although no orthodox comparative static results are possible, by repeated simulation and averaging processes, we may be able to speculate on the comparative static properties of such a model. The importance of the example is two-fold. (*i*) It illustrates how the behavioural theory is able to analyse changes of magnitudes, which the neoclassical theory treats parametrically, as a result of feedback from the environment. (*ii*) It suggests that computer simulation techniques have, to some extent, rendered redundant the argument that the situational information of models must be extremely simple and static if they are to be manageable.

---

[1] Lewin [1935].

[2] Simon [1954], pp. 388–415.

[3] In more sophisticated versions of the model aspirations may be related to 'search' behaviour.

[4] A more dubious candidate for inclusion in the family of behavioural models is Williamson's [1967] managerial model. Three points should be immediately made about this model: (*a*) It is still, unlike Simon's, a maximising model, (*b*) It analyses the decisions of a single large firm, (*c*) The function which the firm maximises is complex and includes such arguments as 'expenditures on staff', 'organisational slack' and 'discretionary investment spending'. Such models lead to what Baumol and Quandt [1964] have called 'optimally imperfect decisions'; optimal because the marginal calculus is applied; imperfect because a superior position could have been reached by the neoclassical 'homo economicus'. Although it is tempting to classify this model within the programme of situational determinism I do not think we should. For unlike situationally deterministic models, Williamson's approach changes the functions to be maximised, or, to put it another way, it introduces complexity in the motivational assumptions of the decision making process. This results from the explicit recognition that Williamson's firm is in a relatively flexible environment where the wish for maximisation is tempered by the existence of conflicting motivations and where survival does not depend on making optimal use of certain objectively defined quantities. Indeed, in his later writings Williamson considers changes in the firm's utility function through very simple learning procedures affecting the firm's behaviour over time. Cf. Williamson [1970], chapter 5.

[5] Cohen and Cyert [1965], pp. 363–76.

These all too sketchy examples of behaviouralism are not intended to demonstrate its success or superiority in dealing with problems in the area of the firm. No doubt, this approach has a potential to deal with problems which neoclassical theory is *a priori* unsuited to. Similarly, neoclassical theory gives some simple answers to questions which we cannot even start asking in terms of behaviouralism (namely, in the domain of market structure and behaviour). It seems to me therefore that questions of appraisal of the rival research programmes is premature, particularly for the budding research programme of behaviouralism.[1] But I can say now that I see no prospect for a compromise solution whereby situational determinism retains most of the theory of the firm but concedes oligopoly to behaviouralism. I have endeavoured to show that the perfectly competitive paradigm entails an all-encompassing method (situational determinism) which has come to be applied to whole range of market situations. In my view behaviouralism, with its emphasis on the psychological characteristics of decision makers, entails a similarly pervasive methodological view.

Thus the two approaches are in my view, importantly different and mutually exclusive over an extensive area. The neoclassical view stresses the *situation* and turns the decision making agent into a cypher; the behavioural view focuses on the nature and characteristics of the *decision*

---

[1] The application of game theory to the problem of oligopoly cannot be subsumed under either situational determinism or behaviouralism although it has common elements with both. In common with behaviouralism it has the recognition that the traditional concept of rationality must be modified in order to deal with multiple exit market situations with decision interdependence. In common with situational determinism it has the view that nevertheless the required psychological assumptions remain extremely weak. Game theory views the competitive situation in terms of the structure of the game matrix which consists of players, strategies and outcomes. The rules of the game permit certain configurations to count as solutions and others not. Equilibrium strategies can be determined in this manner and in the particular case of two-person zero-sum games there are mechanical methods for arriving at solutions. The trouble is that as we move away from two person zero-sum games, psychological and behavioural assumptions (particularly about learning and bargaining behaviour) come increasingly to the forefront. At the same time the techniques of game theory become increasingly complex and awkward to handle as we move away from the simplest games. In the area where the techniques of game theory are most powerful the scope and relevance of the game theoretic approach appears for the moment decidedly limited.

making agent. The distinction between the two is not derived from differences in markets; either can apply to all markets. The claim of 'behaviouralism' is ultimately empirical: even under conditions of apparently severe competition there are systematic patterns of divergence from optimising behaviour.

3 AN APPENDIX ON METHODOLOGICAL CONTROVERSIES: FALSIFICATIONISM, CONVENTIONALISM AND RESEARCH PROGRAMMES IN ECONOMICS

I have identified two research programmes as main rivals in the theory of the firm. I suggested that the neoclassical programme was degenerating. I do not claim that the behaviouralist rival is in much better shape: it does not seem to have successfully predicted any unexpected novel fact, and it seems to deal with various anomalies in a piecemeal, disconnected fashion. Indeed, as a research programme, it is less rich and much less coherent than its neoclassical opponent. One sees a maze of low level generalisations guided mainly by common-sense psychological considerations and growing chaotically in very different directions. No real hard core or comprehensive positive heuristic has emerged yet from the 'behavioural' trial and error.[1] Still, new avenues are worth exploring.

At this point it might be instructive to see what new light the methodology of scientific research programmes sheds on the methodological discussions in economics since the 1930s, especially on those bearing on the theory of the firm. With hindsight, the most interesting conflict was between a Popper-inspired criterion directed against different versions of the neoclassical programme on the one hand, and the apologetic defence of that programme on the other. The most influential figure in the first camp was Hutchison; the second group's leaders were Knight, Mises, Robbins and, later, the Chicago School, and in particular Friedman and Machlup.

(*a*) *Neoclassical Apologetics versus Falsificationism.*

The contributors to this debate have been of two kinds. The main interest of some was philosophical, external to economics proper. They wanted to establish a universal demarcation criterion by which one can sharply distinguish science both from meaningless propositions and from pseudo-scientific ones. They laid down the law and then applied it, come what may. The other sort of contributors were economists who became interested in methodology either in order to defend their favourite theories from a methodological criticism or in order to attack their rivals with methodological arguments.

Hutchison's contribution, in spite of his being a professional economist, was primarily in the first category. His problem was philosophical. It was the demarcation problem; the same problem which the Vienna Circle tried to solve by demarcating tautological, empirical and metaphysical statements. Where does economics belong? If economics is about reality, it ought to be empirical. This was Hutchison's view and his adversaries soon branded him as 'positivist' or 'ultra-empiricist'.

---

[1] There is a third possible rival. Cf. *above*, p. 229, footnote 3.

Hutchison's *The Significance and Basic Postulates of Economic Theory* was published in 1938. He wrote the book in Nazi Germany and his main motivation was to find an appeal-procedure to empirical facts 'against the crank, the propagandist, and the coloured-shirted champion of some persecuting mass creed by whom [he] was surrounded when [he] wrote [his] book.'[1] He wanted to 'mark off statements and theories based on, or deriving from, a disciplined study, from the statements and theories of the undisciplined, pseudoscientific, totalitarian ideologies', the rise of which he regarded as 'the most sinister phenomena of recent decades'.[2] He then came across Popper's *Logik der Forschung* and found in Popper's demarcation criterion the intellectual weapon he wanted.[3] By Popper's criterion Nazi ideology, being unfalsifiable, was to be rejected. But what about neoclassical economics as proposed and defended by Mises or Robbins? Hutchison came to the conclusion that at least a large part of neoclassical economics was also unfalsifiable, and based on—or implying—the less repugnant but equally untestable idea of utopian liberalism.[4]

Hutchison's methodological attack on the neoclassical programme met some furious response from the then leading figure of liberal neoclassical economics, Frank Knight. Hutchison's falsifiability criteria were described by Knight as 'dogmatic' and 'fundamentally misleading and wrong, if not actual nonsense'. Knight, in a lengthy review, condemned Hutchison's 'positivism' and defended neoclassical economics and liberalism with the help of some of the most characteristic tenets of Teutonic '*Wesensschau*', '*verstehende Wissenschaft*', '*Einfühlung*'.[5] In Knight's view neoclassical economic principles were unquestionably true and therefore untestable.[6]

Knight rejected the 'positivist' claim that economics—a '*Geisteswissenschaft*'—should be treated like a natural science: 'I for one, do not particularly care whether economic theory is regarded as a science or not. My concern is that there should be a clear understanding in the minds of students that it is not a science in the sense upon which Hutchison insists in most of his book, namely that of a natural, or especially a physical, science.'[7] Moreover, Knight added, anticipating Oakeshott and Polanyi—there can be no explicitly formulable demarcation criterion even in natural science, for 'agreements on observations and on the principles of logic and their application exist in a community of discourse made up of persons who are (*a*) competent (*b*) trustworthy. There is no 'test' by which such persons may be identified by anyone outside the community of discourse made up of competent and trustworthy judges identifiable only by mutual recognition'.[1] This authoritarianism ruled out any ***articulated*** criterion which—like Popper's—can be applied by outsiders. Good scientific economics is economics done by a good economist.[2]

---

[1] Hutchison [1941], p. 739.

[2] Preface to the reprint of his [1938] ([1960], p. ix).

[3] According to Klappholz and Agassi [1959] it was Hutchison's 'great merit to have introduced Popper's falsifiability criterion into economics'.

[4] Cf. especially his [1937], p. 74 and the *Appendix* of his [1938] and his [1956], p. 483, footnote 8.

[5] Knight [1940], pp. 4, 5, 7, 13, 31; the expressions 'recognition of the essence', 'knowledge based on understanding', 'sympathetic introspection', are technical terms of the German hermeneutic philosophy which influenced Knight. It is interesting to note that much of the '*Positivismusstreit*' of the late 1960s in Germany was anticipated in the Knight-Hutchison debate. Incidentally, Knight anticipates Kuhn's idea of 'testing' within 'normal science': 'Any tests which can be proposed . . . have to be tested by the propositions of economic theory as already understood' (Knight [1940], p. 3).

[6] A. W. Stonier also wrote a briefly contemptuous review of Hutchison's 'application of logical positivism to economic theory' (Stonier [1939]).

[7] Knight [1941], p. 751. It is interesting to note that twenty years earlier Knight wrote: 'In fact economics is different from physics in degree, since, though it cannot well be made so exact, yet for special reasons it secures a moderate degree of exactness only at the cost of much greater unreality' ([1921], p. 3).

Let us now turn from Knight's antinaturalism and authoritarianism to some apologetic methodologies, more in keeping with the Anglo-Saxon 'positivist' tradition, put forward in the forties and fifties primarily by Friedman and Machlup.[3] They—unlike Knight—do not wish to apply peculiar standards to the social sciences. They wanted what I call the neoclassical research programme to come out as satisfactory by general methodological standards; at least, it should not be 'impatiently' rejected at the behest of utopian norms. The crux of the matter was that while falsificationists demand that the scientist should specify in advance potential empirical falsifiers, not just for some specific version of his general theory but for his fundamental assumptions, neoclassical economists were unable and unwilling to specify such potential falsifiers for the hard core of their programme, and they wanted to justify their attitude. They also tried to argue that the unit of intellectual appraisal should not be an isolated ***hypothesis*** but rather—though they did not use this term—a research programme. I shall try to show that most of the obscurities in this methodological debate disappear once we regard it as a defence of a research programme against 'naive falsificationism'.

The most important retort to Hutchison's naive falsificationism is to be found in Machlup's [1955] where he unfolds a lucid and in many ways convincing critique of '*Ultra-Empirical Economics*'. Ultraempiricism, for him, is the doctrine which 'refuses to recognise the legitimacy of employing at any level of analysis, propositions not independently verifiable'[4]: ultraempiricists want each single proposition in their system to be 'independently' testable. He readily admits that the 'fundamental assumptions' of neoclassical economics cannot be tested in isolation[5]; he thinks, following Knight, that they only need be 'understandable'[6]: 'We need not be worried about ***independent*** verifications of the fundamental assumptions.'[7] 'Verification is only a must for strictly empirical hypotheses.'[8]

Hutchison, in a reply to Machlup, denies the charge that he wants each hypothesis to be ***directly*** testable in isolation. He only wanted each hypothesis to be '***indirectly***' testable in isolation, that is, ***each*** hypothesis should 'be reducible ***by direct deduction*** to an empirically testable proposition or propositions'.[9] Machlup, in his cogent reply, points out that this is still an unreasonable demand. It is worth quoting Machlup in full:

'Do we really mean the same thing when we speak of "indirect testing"? Perhaps the crucial misunderstanding lies right here. Professor Hutchison mentions that a proposition not itself empirically testable directly must be "reducible by ***direct*** deduction to an empirically testable proposition or propositions". This formulation suggests a requirement that the implications of any single proposition be tested independently of those of other propositions with which it is conjoined to constitute a "case". In fact, however, ***the conjunction of logically independent propositions and derivation of their joint consequences*** is the essence of indirect testing.

---

[1] *Ibid.*, p. 751.

[2] For a half-critical discussion of this 'elitist-authoritarian' position, cf. Lakatos [1971*a*], pp. 120–2.

[3] Friedman [1953] and Machlup [1952] and [1955].

[4] Machlup [1955], p. 7.

[5] Machlup [1955], p. 16.

[6] *Ibid.*, p. 17.

[7] *Ibid.*, my italics.

[8] *Ibid.*, p. 197.

[9] Hutchison [1956], pp. 476–7; my italics.

'If assumption *A* can neither be subjected to any direct empirical test nor reduced "by direct deduction to an empirically testable proposition", its indirect verification can be accomplished by combining it with an assumption *B* which is directly testable; if a consequence *C* can be deduced from the conjunctive hypothesis *A plus B*—but not from either one alone—and if *C* is empirically tested, *A* is regarded as having passed the indirect test.

'I suspect that Professor Hutchison does not accept the validity of indirect verification in this sense.'[1]

Of course Hutchison cannot; for should he accept it, the fundamental assumptions of neoclassical economics would cease to be 'tautological' but would become empirically (even if indirectly) testable. Machlup, on the other hand, can cheerfully accept the validity of indirect testability ***in this sense***, since it vindicates his Basic Assumption:

'Our discussion of the "kind of test" to which the Assumption should be subjected has probably been sufficient to warrant our conclusion that the test of the pudding lies in the eating and not in its ingredients. If we find no better theory to explain or predict changes in prices, outputs, *etc.*, *etc.*, and if our present theory does no worse than it has been doing, we may consider our Assumption as warranted'.[2]

Hutchison's confusion may have been due to the fact that in Popper's *Logik der Forschung*, the concept of 'background knowledge' is not explicitly discussed: this notion is dealt with in Popper's *Conjectures and Refutations*. In the latter, Popper uses 'requirement of independent testability' in two very different senses; in fact he has two very different requirements each called by him by the same name: first he calls the requirement of independent testability that a good theory should explain more than its predecessor and that in general any explanation worthy of its name should say something more than the original explanandum.[3] On the other hand, he also expects (opposing the Duhem-Quine thesis) that the premises of a theory will be independently testable in a very different sense, namely that the guilty premise may, under favourable circumstances, be if only tentatively, identified: 'it is possible in quite a few cases to find which hypothesis is responsible for the refutation of a cluster of hypotheses.'[4] (This is very different from the previous requirement and is identified and criticised by Lakatos as the fourth type of ***decision*** needed by ***naive*** falsificationism.[5]) Machlup's criticism of Hutchison is an almost complete anticipation of Lakatos's similar criticism.

Machlup's final assault on Hutchison is a heuristic one: 'If the ultraempiricists reject . . . any theoretical system that is built on unverifiable or unverified assumptions what is the alternative heuristic they offer? A programme that begins with facts rather than assumptions.'[1] Machlup then quotes Hutchison's unfortunate slip into the heuristic suggestion to abandon abstract theory and replace it 'by statistical investigations, questionnaires to consumers and entrepreneurs, the examination of family budgets and the like'. And Machlup correctly points out that Hutchison's low-level trial-and-error method gives no guidance to statistical fact-collection and gives no answer to the questions: 'Precisely what data would be obtained and statistically investigated? What questions asked of consumers and entrepreneurs?'[2]

---

[1] Machlup [1956], p. 484; my italics.

[2] Machlup [1956], p. 489.

[3] Cf. Popper [1962], p. 244.

[4] Popper [1962], p. 239. Also cf. his [1957], p. 132.

[5] Lakatos [1970], p. 110; for Lakatos's way out from the Duhem-Quine situations see his [1970], pp. 184–8.

From the point of view of the methodology of research programmes it is clear what Machlup is really after. If one rules out hard cores of major research programmes as unscientific one may well rule out Newton's laws and the fundamental assumptions of neoclassical economics[3]: and all one is left with are low-level generalisations like 'all swans are white' which are separately testable.

How can one reconcile research programmes with the criterion of high empirical content? In Lakatos's approach the hard core in isolation has no potential falsifiers, though the various 'falsifiable versions' of the programme do. Machlup comes very near to Lakatos's distinction of ***hard core*** and ***protective belt*** when he compares a theory with a 'machine' with 'fixed parts' and 'exchangeable parts'.[4] Indeed, Machlup ends up with a suggestion of appraisal which comes again near to Lakatos's '***progressive problem-shift***': 'The various successive estimates of ***changeable structures***, [*i.e.* predictions in the falsifiable versions of hard core plus protective belt plus initial conditions], may then be regarded as verifications of general hypotheses [*i.e.* the hard core.]'[5] (There is, however, nothing in Machlup's text about predictions of novel facts.) When however, it comes to the defence of the fundamental neoclassical assumptions from naive falsificationism, he gives something like a 'sophisticated elimination rule': 'Under what circumstances is the Assumption to be regarded as "disconfirmed?" When a theory not using this Assumption is proposed and is shown to work equally well for a wider range of problems, or with a smaller number of variables or provisos, or more reliably or more accurately for the same range of problems and with the same number of variables or qualifications—then the Assumption will have outlived its usefulness and will be sent to the limbo of "disconfirmed propositions". And even this need not be beyond recall.'[6] That the Assumption certainly 'does not always conform to the facts' is not sufficient reason to give it up as part of the hard core of a research programme.[7]

(*b*) *A Falsificationist Plea for Metaphysical Statements in Economics.*

A consideration of Klappholz's and Agassi's critical comment on Hutchison's book[1] may be helpful for a further elucidation of the role of the falsifiability criterion in the methodology of economics.

---

[1] Machlup [1955], p. 8.

[2] *Ibid.*, p. 8, footnote 25.

[3] Machlup points out that Newton's laws may also be regarded as 'definitions': if neoclassical economics is dismissed as unscientific on Hutchisonian grounds, so is Newton. (*Ibid.*, p. 16.) Cf. also *below*, p. 240.

[4] 'The so called *fundamental assumptions* are fixed part of the machine; they make the machine what it is; they cannot be changed without changing the character of the entire machine. All the other parts are exchangeable like coils, relay spools, wires, tapes, cylinders, records or mats, something that can be selected and put in, and again taken out to be replaced by a different piece of the set'. (*Ibid.*, p. 12.)

[5] *Ibid.*, p. 21.

[6] Machlup [1956], p. 489. It should be mentioned in this connection that Machlup's [1952] is a fascinating though intricate attempt to accommodate new developments in the theory of the firm within the neoclassical research programme.

[7] *Ibid.*, p. 488.

Klappholz and Agassi make two points. This first is that Hutchison misread Popper's falsificationism and in doing so committed a grave mistake. They accuse Hutchison of having drawn a logically invalid dichotomy between 'tautologous' and 'empirical'.[2] There is, they claim, a third region: metaphysics. Hutchison, according to Klappholz and Agassi, conflated 'conceivably false' with 'conceivably refutable'. But metaphysical statements—and here they refer to Popper—are conceivably false but not conceivably refutable.[3] The dichotomy has to be replaced by a trichotomy.

Hutchison retorts that for all the 'obscurity' of 'conceivable falsifiability', surely *all conceivably false statements are conceivably refutable by observation*[4]: and then, if 'empirical' means 'conceivably refutable' or 'refutable in principle', his dichotomy holds.

It is worthwhile to sort out the rights and wrongs of this controversy 'Testability' in Popper's texts seems normally to refer to testability (that is, to availability of potential falsifiers) at some given time *t*. If so, a proposition *p* might have been untestable in Popper's sense in 1600 but testable in 1900 (for example, the existence of stellar parallax). The 'material requirement' for a proposition to be a potential falsifier depends on the experimental techniques of the time.[5] Clearly, in Popper's sense some factual, contingent propositions are testable, some metaphysical.[6] But are not *all* factual propositions '*in principle*' testable? Suppose we adopted the following plausible interpretation of 'testability in principle': a proposition *p* is in principle testable if additional propositions (expressing initial conditions or suitable observational theories) can conceivably be devised such that: (*1*) The conjunction of these additional propositions, call them *b*, and *p* is consistent and has potential falsifiers; (*2*) The additional propositions can be accepted as belonging to 'background knowledge' and (*3*) the conjunction of *p* and *b* can be refutable by some proposition which can be accepted, by convention, as an 'observation' statement.[7]

Now it is interesting that if one follows this definition through the Popper-Watkinsian trichotomy disappears for *conceivable falsifiability*. For once we bring in background knowledge into the falsifying hypothesis, all-some statements and even purely existential statements become conceivably falsifiable since both *might* be falsified (and therefore *are*, 'conceivably falsifiable') by accepted background knowledge which, of course, contains universal statements: and a universal statement may well clash with pure existential and all-some statements. For instance, 'all metals have a solvent' is conceivably falsifiable with the help of a well-corroborated theory h_1 which, through a stability theory of nuclei, restricts the number of metals to a finite number and with the help of some classical theory h_2 similarly restricting the number of solvents.

---

[1] Klappholz and Agassi [1959] and [1960].

[2] 'Hutchison's dichotomy is false'. (*Ibid.*, p. 65.)

[3] 'Metaphysical statements forbid some conceivable occurrence and yet are not conceivably falsifiable by observation' (*loc. cit.*, p. 63). This theme had been elaborated in Watkins [1957] and [1958].

[4] Hutchison [1960], p. xxi, footnote 1.

[5] Cf. Lakatos [1968*a*], p. 378.

[6] Cf. especially Watkins [1957] and [1958].

[7] For the Popperian distinction between observational and convention-laden 'observational', cf. Lakatos [1970], p. 107. Only if 'by observation' means 'with the help of some conceivable theories and/or techniques', might Hutchison's dichotomy be saved.

This consideration shows how the explicit introduction of 'background knowledge' raises new problems for falsificationists. However, I do not want to claim that Hutchison was aware of these distinctions. Indeed, Klappholz and Agassi were right in pointing out that Hutchison's actual examples (like the principle of profit maximisation, or other economic principles with implicit or explicit *ceteris paribus* clauses) were, syntactically speaking, not tautologies: they were conceivably false, factual, contingent statements. Yet they were unfair to the spirit of Hutchison's text: in most places he did not claim that they were tautologies but only claimed that they were treated *as if* they were tautologies,[1] *i.e.* treated as if they were unfalsifiable sentences: *i.e.* high level conventions that are to be taken as unfalsifiable. In this interpretation it is merely a semantical question whether to call a proposition which we have decided to make irrefutable, 'tautological' or 'metaphysical'.[2] One suspects that Hutchison might have stumbled on the phenomenon of the 'hard core'; but he conflated convention with tautology. But his critics showed his confusion without disentangling its rationale; that is, without admitting that the fundamental assumptions of neo-classical economics are not just 'not yet testable' but rather 'not testable any more'. These assumptions have been hardened into a hard core of a major programme and are now protected by a protective belt. This hard' core' functions as a convention although formally it is synthetic; its *heuristic role* empties it of 'empirical content' (*i.e.* of potential falsifiers). This gave to Hutchison and to many others the illusion that the hard core was a tautology in the logical sense.[3]

(It is interesting that when Popper himself came to discuss his Rationality Principle, he also suggested that it was to be irrefutable by methodological decision and devoid of empirical content in this sense. Such hard cores can, then, never be displaced by empirical counter-evidence but only by a—theoretically and empirically—more progressive research programme. Thus Popper's schema of explanation in the social sciences is out of step with his falsifiability criterion but may be clarified in the light of the methodology of research programmes.)

This leads to the second strand of Klappholz's and Agassi's argument and the one which they regard as 'primary'[4]: That Hutchison's falsifiability criterion showed 'an attitude of impatience and haste . . . prevalent in methodological writings in economics'; that Hutchison's dichotomy 'justified the impatient dismissal of ideas which are not yet testable (but which may become testable by further specification)'.[1]

---

[1] Cf. Hutchison [1960], p. 158.

[2] Agassi in his [1971] changed his mind and followed Hutchison into the confusion: 'However idiosyncratic a convention is when it is consistently applied it is a tautology' (Agassi [1971], p. 62). But does a metaphysical principle, if methodologically treated as a convention become a tautology? Yes, if Agassi regards metaphysical statements as implicit definitions. But then in Agassi's [1971] view Hutchison's [1938] dichotomy (contrary to Agassi's [1959] stand) must hold. (Incidentally, Agassi borrows some of his concrete examples and his conclusion from Klappholz and Mishan [1962] although Klappholz's and Mishan's paper is more clearly worded.)

[3] Popper himself was fully aware of the important role of conventions in scientific method. It may be worthwhile to quote Popper in full: '... it is impossible to decide, by analysing its logical form, whether a system of statements is a conventional system of irrefutable implicit definitions, or whether it is a system which is empirical in my sense; that is, a refutable system. This however only goes to show that my criterion of demarcation cannot be applied immediately to a *system of statements*.... The question whether a given *system* should as such be regarded as a conventionalist or an empirical one is therefore misconceived. *Only with reference to the method applied* to a theoretical system is it at all possible to ask whether we are dealing with a conventionalist or an empirical theory' (Popper [1934], section 20).

[4] Klappholz and Agassi [1960], p. 160.

A problem remains: If we do not dismiss untestable propositions as unscientific what propositions *do* we dismiss as unscientific? Klappholz and Agassi give no reply.[2]

(*c*) *The 'F-twist' versus Chamberlinian 'Realism'.*

The criticism of the Chicago version of neoclassical economics was pursued along two lines. The one which we already have discussed was that its fundamental assumptions were obviously false: one does not need much fieldwork to discover that not all firms are intent on profit maximisation.

The Chicago School's self defence may be easily reformulated in Lakatos's terminology: First they pointed out that anomalies need not deprive a research programme of its progressive character. Secondly, even obvious falsity of the hard core in the standard interpretation need not deprive a research programme of its progressive character. Remember that Newton himself believed that action at a distance was an absurd, blatantly false idea; but this did not prevent the triumphant progress of his research programme; only a problem of reduction remained open, namely the problem of explaining why planets behave *as if* they attracted each other.[3] In Friedman's case, one would have to explain why the market behaves *as if* firms wanted to maximise profit *etc.*[4] These unsolved problems of reduction of the hard core to 'more realistic' ones do not prevent either the theoretical or the empirical progress of the programme. Empirical falsifications coming from 'indirect' tests do not harm progress either. Anomalies, unsolved puzzles can always abound: they only provide economists with work in elaborating the protective belt.[5] So far Friedman's position is vindicated by

---

[1] *Ibid.*

[2] Watkins in his [1957] puts the problem as follows: 'Since influential haunted-universe doctrines (which permeate science) are neither demonstrable nor testable, it becomes urgent to investigate ways in which they can be rationally supported and criticised' (Watkins [1957], p. 131 and [1958], pp. 363–5). Unfortunately, this problem was not understood in the neighbouring economics department. In 1963 Lipsey published the first edition of his influential textbook (Lipsey [1963]), which can be fairly described as dogmatic falsificationist: it claimed that theories cannot be proved but can be disproved. By the time of the second edition Lipsey discovered that 'it is also impossible to refute any theory conclusively (Lipsey [1966], p. 16) and therefore announced that he abandoned the Popperian notion of refutation (Lipsey [1966], p. 52) and shifted to probabilism. Archibald at roughly the same time also, shifted to probabilism but in this case from a more sophisticated falsificationist position (Archibald [1961]). For a Popperian criticism of Chamberlin's theory of monopolistic competition for *ad hocness*, cf. Archibald [1961] and also Lipsey [1963], p. 220 especially footnote 1.

[3] This problem is connected with whether one interprets the hard core of a programme in an 'essentialist' or in a 'nominalist' way (cf. Lakatos [1970], p. 126). Essentialism corresponds in the Chicago-Chamberlin debate to 'realism' and nominalism to 'conventionalism' or 'instrumentalism'. Also cf. Lakatos, *op. cit.*, p. 188, footnote 3.

[4] Friedman [1953], pp. 21 ff.; for an unexciting criticism, cf. Melitz [1965], pp. 45–51.

[5] Klappholz and Agassi in their [1959] misinterpreted Friedman when they claimed that according to Friedman the Chicago approach is simply unrefuted, in a Popperian sense, and therefore 'acceptable' (p. 68). What Friedman says in this [1953]—if interpreted in the context of his whole methodology—is that anomalies can be solved within the

Lakatos's methodology. False assumptions may be rich in true (and also in false) consequences, and long term progress may be founded on an intuitively false hard core. The trouble comes only when we appraise the theoretical and empirical progress of the programme: the *ad hocness* or *non-ad hocness* of the adjustments in the ever more complicated protective belt. This is, as I tried to show, where the neoclassical programme—after a period of initial progress—started failing. Friedman never seriously analysed the theoretical adjustments made by the neoclassical school (which he openly defends) for *ad hocness*.

Thus the Friedman-Machlup methodology can easily be characterised as an attempt to defend a research programme against utopian standards of falsificationism.[1] Their only major slip in these defensive manoeuvres was Friedman's much discussed thesis—christened by Samuelson the 'F-twist'[2]—that 'to be important . . . a hypothesis must be descriptively false in its assumptions'.[3] This was nothing more than a blunt and provocative formulation of the truism that the hard core of a powerful research programme may consist of counter-intuitive over-simplifications. The real weakness of this methodology is its purely defensive character and its lack of a clear empirical criterion of progress.[4]

Chicago approach; he does not deny that at any given time there are open specific concrete problems (p. 23) the solving of which puts (in a Kuhnian-Lakatosian sense) rather the economist than the economic theory to test. Indeed, Friedman says that what distinguishes a 'professional' from a 'crackpot' is the ability to distinguish a serious counter-example from a mere anomaly or from some phenomenon which does not come at all under the purview of the programme (p. 25).

[1] Classical inductivism—the requirement of proving one's theories from facts—has rarely been seriously presented in recent debates. The only such attempt occurred in one or two of Samuelson's too many contributions to the methodological debate in the 1950s and the 1960s in the columns of the *American Economic Review*. (Samuelson [1952], p. 57 and [1963], pp. 231–6; but see Machlup [1964], pp. 733–5.)

[2] Samuelson [1963], p. 232.

[3] Friedman [1953], p. 15.

[4] (*Added in the press*): Recently Kuhnians who were first to show up some of the weaknesses of what they christened 'naive falsificationism' came to the fore of methodological discussion also in economics to save 'paradigms' from falsificationists; cf. *e.g.* Coats [1972] and Loashby [1971]. They appeared in economics at a time when Kuhn himself withdrew or weakened most of his sharp theses (cf. *e.g.* Musgrave [1971]). But does not the methodology of research programmes retain all the advantages of the Kuhnian approach without its disadvantages?

Department of Philosophy, Logic and Scientific Method
London School of Economics

REFERENCES

Agassi, J. [1971]: 'Tautology and Testability in Economics', *Philosophy of the Social Sciences*, 1, pp. 49–63.

Anscombe, E. [1957]: *Intention*.

Archibald, G. C. [1961]: 'Chamberlin *versus* Chicago', *Review of Economic Studies*, 29, pp. 2–21.

Archibald, G. C. [1965]: 'The Qualitative Content of Maximising Models', *The Journal of Political Economy*, 73, pp. 27–36.

Bain, J. S. [1949]: 'A Note on Pricing in Monopoly and Oligopoly', *The American Economic Review*, 39, pp. 448–64.

Bain, J. S. [1956]: *Barriers to New Competition*.

Bain, J. S. [1967]: 'Chamberlin's Impact on Microeconomic Theory', in R. E. Kuenne (*ed.*): *Monopolistic Competition Theory: Studies in Impact*, pp. 147–76.

BAUER, P. T. [1945]: 'Notes on Cost', *Economica (New Series)*, **12**, pp. 90–100.

BAUMOL, W. J. [1965]: *Economic Theory and Operations Analysis.*

BAUMOL, W. J. and QUANDT, R. E. [1964]: 'Rules of Thumb and Optimally Imperfect Decisions', *The American Economic Review*, **54**, pp. 23–46.

BAUMOL, W. J. and STEWART, M. [1971]: 'On the Behavioural Theory of the Firm', in R. M. Marris and A. Wood (*eds.*): *The Corporate Economy*, pp. 118–43.

BERTRAND, J. [1883]: Review of Cournot's *Researches*, *Journal des Savants*, September, 1883.

BISHOP, R. L. [1964]: 'The Theory of Imperfect Competition after Thirty Years: The Impact on General Theory', *The American Economic Review*, **54**, pp. 33–43.

CHAMBERLIN, E. H. [1933]: *The Theory of Monopolistic Competition.*

CHAMBERLIN, E. H. [1957]: *Towards a More General Theory of Value.*

CHAMBERLIN, E. H. [1966]: *The Theory of Monopolistic Competition* (*Sixth Edition*).

CLAPHAM, J. H. [1922]: 'Of Empty Economic Boxes', *Economic Journal*, **32**, pp. 305–14.

CLARK, J. B. [1907]: *Essential of Economic Theory.*

COATS, A. W. [1972]: 'The Economic and Social Context of the Marginal Revolution of the 1870's', *History of Political Economy*, **4**. Reprinted in C. Goodwin, R. D. C. Black and A. W. Coats (eds.): *The Marginal Revolution.*

COHEN, K. J. and CYERT, R. M. [1965]: *Theory of the Firm.*

COURNOT, A. A. [1838]: *Récherches sur les Principes Mathématiques de la théorie des richesses.*

CYERT, R. M. and MARCH, J. G. [1963]: *Behavioral Theory of the Firm.*

DAVIDSON, D. [1963]: 'Actions, Reasons and Causes', *Journal of Philosophy*, **60**, pp. 685–700.

DEMSETZ, H. [1964]: 'The Welfare and Empirical Implications of Monopolistic Competition Theory', *Journal of Political Economy*, **74**, pp. 623–41.

DEUTSCH, M. and KRAUSS, R. M. [1960]: 'The Effect of Threat upon Inter-personal Bargaining', *Journal of Abnormal and Social Psychology*, **61**, pp. 181–89.

DEWEY, D. [1969]: *The Theory of Imperfect Competition.*

DONAGAN, A. [1970]: 'Comment', in R. Borger and F. Cioffi (*eds.*): *Explanations in the Behavioural Sciences*, pp. 218–227.

ENKE, S. [1951]: 'On Maximising Profits', *The American Economic Review*, **41**, pp. 566–78.

FELLNER, W. [1949]: *Competition Among the Few.*

FRIEDMAN, M. [1953]: *Essays in Positive Economics.*

HEBB, D. O. [1949]: *Organisation of Behavior.*

HEBB, D. O. [1955]: 'Drives and the Conceptual Nervous System', *Psychological Review*, **62**, pp. 243–54.

HEMPEL, C. G. [1965]: *Aspects of Scientific Explanation.*

HICKS, J. R. [1939]: *Value and Capital.*

HIGGINS, B. [1939]: 'Elements of Indeterminacy in the Theory of Non-Perfect Competition', *The American Economic Review*, **29**, pp. 468–79.

HUTCHISON, T. W. [1937]: 'Note on Uncertainty and Planning', *The Review of Economic Studies*, **5**, pp. 72–74.

HUTCHISON, T. W. [1938]: *The Significance and Basic Postulates of Economic Theory.*

HUTCHISON, T. W. [1941]: 'The Significance and Basic Postulates of Economic Theory: A Reply to Professor Knight', *The Journal of Political Economy*, **19**, pp. 732–50.

HUTCHISON, T. W. [1956]: 'Professor Machlup on Verification in Economics', *The Southern Economic Journal*, **22**, pp. 476–83.

HUTCHISON, T. W. [1960]: *The Significance and Basic Postulates of Economic Theory* (Second edition).

JAMES, W. [1891]: *The Principles of Psychology.*

JOHNSON, H. L. and COHEN, A. M. [1967]: 'Experiments in Behavioral Economics: Siegel and Fouraker Revisited', *Behavioral Science*, **12**, pp. 353–72.

KALDOR, N. [1935]: 'Market Imperfection and Excess Capacity', *Economica (New Series)*, **2**, pp. 33–50.

KALDOR, N. [1938]: 'Professor Chamberlin on Monopolistic and Imperfect Competition', *Quarterly Journal of Economics*, **52**, pp. 513–38.

KELLEY, H. H. [1965]: 'Experimental Studies of Threats in Interpersonal Negotiations', *Journal of Conflict Resolution*, **9**, pp. 79–105.

KLAPPHOLZ, K. and AGASSI, J. [1959]: 'Methodological Prescriptions in Economics', *Economica (New Series)*, **26**, pp. 60–74.

Bauer, P. T. [1945]: 'Notes on Cost', *Economica (New Series)*, **12**, pp. 90–100.

Baumol, W. J. [1965]: *Economic Theory and Operations Analysis.*

Baumol, W. J. and Quandt, R. E. [1964]: 'Rules of Thumb and Optimally Imperfect Decisions', *The American Economic Review*, **54**, pp. 23–46.

Baumol, W. J. and Stewart, M. [1971]: 'On the Behavioural Theory of the Firm', in R. M. Marris and A. Wood (*eds.*): *The Corporate Economy*, pp. 118–43.

Bertrand, J. [1883]: Review of Cournot's *Researches*, *Journal des Savants*, September, 1883.

Bishop, R. L. [1964]: 'The Theory of Imperfect Competition after Thirty Years: The Impact on General Theory', *The American Economic Review*, **54**, pp. 33–43.

Chamberlin, E. H. [1933]: *The Theory of Monopolistic Competition.*

Chamberlin, E. H. [1957]: *Towards a More General Theory of Value.*

Chamberlin, E. H. [1966]: *The Theory of Monopolistic Competition (Sixth Edition).*

Clapham, J. H. [1922]: 'Of Empty Economic Boxes', *Economic Journal*, **32**, pp. 305–14.

Clark, J. B. [1907]: *Essential of Economic Theory.*

Coats, A. W. [1972]: 'The Economic and Social Context of the Marginal Revolution of the 1870's', *History of Political Economy*, **4**. Reprinted in C. Goodwin, R. D. C. Black and A. W. Coats (eds.): *The Marginal Revolution.*

Cohen, K. J. and Cyert, R. M. [1965]: *Theory of the Firm.*

Cournot, A. A. [1838]: *Récherches sur les Principes Mathématiques de la théorie des richesses.*

Cyert, R. M. and March, J. G. [1963]: *Behavioral Theory of the Firm.*

Davidson, D. [1963]: 'Actions, Reasons and Causes', *Journal of Philosophy*, **60**, pp. 685–700.

Demsetz, H. [1964]: 'The Welfare and Empirical Implications of Monopolistic Competition Theory', *Journal of Political Economy*, **74**, pp. 623–41.

Deutsch, M. and Krauss, R. M. [1960]: 'The Effect of Threat upon Inter-personal Bargaining', *Journal of Abnormal and Social Psychology*, **61**, pp. 181–89.

Dewey, D. [1969]: *The Theory of Imperfect Competition.*

Donagan, A. [1970]: 'Comment', in R. Borger and F. Cioffi (*eds.*): *Explanations in the Behavioural Sciences*, pp. 218–227.

Enke, S. [1951]: 'On Maximising Profits', *The American Economic Review*, **41**, pp. 566–78.

Fellner, W. [1949]: *Competition Among the Few.*

Friedman, M. [1953]: *Essays in Positive Economics.*

Hebb, D. O. [1949]: *Organisation of Behavior.*

Hebb, D. O. [1955]: 'Drives and the Conceptual Nervous System', *Psychological Review*, **62**, pp. 243–54.

Hempel, C. G. [1965]: *Aspects of Scientific Explanation.*

Hicks, J. R. [1939]: *Value and Capital.*

Higgins, B. [1939]: 'Elements of Indeterminacy in the Theory of Non-Perfect Competition', *The American Economic Review*, **29**, pp. 468–79.

Hutchison, T. W. [1937]: 'Note on Uncertainty and Planning', *The Review of Economic Studies*, **5**, pp. 72–74.

Hutchison, T. W. [1938]: *The Significance and Basic Postulates of Economic Theory.*

Hutchison, T. W. [1941]: 'The Significance and Basic Postulates of Economic Theory: A Reply to Professor Knight', *The Journal of Political Economy*, **19**, pp. 732–50.

Hutchison, T. W. [1956]: 'Professor Machlup on Verification in Economics', *The Southern Economic Journal*, **22**, pp. 476–83.

Hutchison, T. W. [1960]: *The Significance and Basic Postulates of Economic Theory* (Second edition).

James, W. [1891]: *The Principles of Psychology.*

Johnson, H. L. and Cohen, A. M. [1967]: 'Experiments in Behavioral Economics: Siegel and Fouraker Revisited', *Behavioral Science*, **12**, pp. 353–72.

Kaldor, N. [1935]: 'Market Imperfection and Excess Capacity', *Economica (New Series)*, **2**, pp. 33–50.

Kaldor, N. [1938]: 'Professor Chamberlin on Monopolistic and Imperfect Competition', *Quarterly Journal of Economics*, **52**, pp. 513–38.

Kelley, H. H. [1965]: 'Experimental Studies of Threats in Interpersonal Negotiations', *Journal of Conflict Resolution*, **9**, pp. 79–105.

Klappholz, K. and Agassi, J. [1959]: 'Methodological Prescriptions in Economics', *Economica (New Series)*, **26**, pp. 60–74.

Popper, K. R. [1957]: *The Poverty of Historicism.*
Popper, K. R. [1959]: *The Logic of Scientific Discovery.*
Popper, K. R. [1962]: *Conjectures and Refutations.*
Popper, K. R. [1967]: 'La Rationalité et le statut du Principe de Rationalité', in E. M. Claassen (*ed.*): *Le Fondements Philosophiques des Systèmes Économiques: Textes de Jacques Rueff et Essais Rédigés en son Honneur*, pp. 142–50.
Robinson, R. [1971]: *Edward H. Chamberlin.*
Rothchild, K. [1947]: 'Price Theory and Oligopoly', *Economic Journal*, **57**, pp. 299–320.
Ryle, G. [1949]: *The Concept of Mind.*
Samuelson, P. A. [1947]: *Foundations of Economic Analysis.*
Samuelson, P. A. [1952]: 'Economic Theory and Mathematics—An Appraisal', *American Economic Review*, **42**, Number 2, pp. 55–66.
Samuelson, P. A. [1963]: Comment on Ernest Nagel: 'Assumptions In Economic Theory', *The American Economic Review*, **53**, pp. 231–36.
Samuelson, P. A. [1967]: 'The Monopolistic Competition Revolution', in R. E. Kuenne (*ed.*): *Monopolistic Competition Theory: Studies in Impact*, pp. 105–38.
Selznick, P. [1949]: *TVA and the Grass Roots.*
Simon, H. A. [1945]: *Administrative Behavior.*
Simon, H. A. [1954]: 'Some Strategic Considerations in the Construction of Social Science Models'. in P. Lazarsfeld (*ed.*): *Mathematical Thinking in the Social Sciences*, pp. 389–415.
Simon, H. A. [1955]: 'A Behavioural Model of Rational Choice', *The Quarterly Journal of Economics*, **69**, pp. 99–118.
Simon, H. A. [1957]: *Models of Man.*
Simon, H. A. [1960]: *The New Science of Management Decision.*
Simon, H. A. [1967]: 'Motivational and Emotional Controls to Cognition', *Psychological Review*, **74**, pp. 29–39.
Simon, H. A. [1968]: 'On Judging the Plausibility of Theories', in B. Van Rootselaar and J. F. Staal (*eds.*): *Logic, Methodology and Philosophy of Science*, pp. 439–62.
Simon, H. A. [1972]: 'Theories of Bounded Rationality', in C. B. McGuire and R. Radner (*eds.*) *Decision and Organisation*, pp. 101–30.
Smith, A. [1976]: *The Wealth of Nations.*
Sraffa, P. [1926]: 'The Laws of Returns under Competitive Conditions', *Economic Journal*, **36**, pp. 535–50.
Stackelberg, H. [1934]: *Marktform und Gleichgewicht.*
Stigler, G. J. [1949]: 'Monopolistic Competition in Retrospect', in his *Five Lectures on Economic Problems.*
Stigler, G. J. [1957]: 'Perfect Competition Historically Contemplated', *The Journal of Political Economy*, **65**, pp. 1–16.
Stigler, G. J. [1966]: *The Theory of Price* (*Third edition*).
Stonier, A. W. [1939]: Review of T. W. Hutchison: *The Significance and Basic Postulates of Economic Theory, The Economic Journal*, **49**, pp. 114–5.
Sylos-Labini, P. [1956]: *Oligopolio e Progresso Technico.*
Thibaut, J. S. and Kelley, H. H. [1959]: *The Social Psychology of Groups.*
Thirlby, G. F. [1952]: 'The Economist's Description of Business Behavior', *Economica*, **19**, pp. 148–167.
Triffin, R. [1940]: *Monopolistic Competition and General Equilibrium Theory.*
Watkins, J. W. N. [1957]: 'Between Analytic and Empirical', *Philosophy*, **32**, pp. 112–31.
Watkins, J. W. N. [1958]: 'Influential and Confirmable Metaphysics', *Mind*, **67**, pp. 344–65.
Watkins, J. W. N. [1970]: 'Imperfect Rationality', in R. Borger and F. Cioffi (*eds.*): *Explanations in the Behavioural Sciences*, pp. 167–217.
Watkins, J. W. N. [1970]: 'Reply', in R. Borger and F. Cioffi (*eds.*): *Explanations in the Behavioural Sciences*, pp. 228–30.
Williamson, O. E. [1967]: *The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm.*
Williamson, O. E. [1970]: *Corporate Control and Business Behavior.*

Discussions

METAPHYSICAL COMMON SENSE

Professor Kneale, in three recent articles ([1971], [1972*a*], [1972*b*]), has defended three distinctions which he holds to be analogous to one another, between propositions and sentences, between predicates and predicative phrases, and between numbers and numerals. Propositions, predicates, and numbers are *kinds* to which utterances may belong; a token of a sentence or propositional phrase is a specimen of a proposition when it presents a possible state of affairs for consideration, a token of a predicative phrase is a specimen of a predicate when it presents a property that things may have, and a token of a numeral is a specimen of a number when it presents a possible 'quotity' or 'extent of exemplification'. Various utterances will belong to one of *these* kinds in virtue of a shared role in communication, whereas they will be classified differently on other grounds as specimens of the same sentence, *etc.*

Kneale uses these distinctions for several purposes; among others, to solve a number of paradoxes, to give ([1972*a*], Part IV) a definition of truth which, in words, would be

'X is true' $=_{def.}$ 'For some p, X is the-proposition-that-p and p'

and to defend, against Tarski, the consistency and importance of natural languages in which we can talk without hierarchical restrictions about the expressions of those languages, about their meaning, truth, and falsity, and about propositions, predicates, and so on.

I find myself in very full agreement with Kneale's approach. Distinctions which we undoubtedly need can be drawn in the way he suggests. (I am not quite sure that numbers, as he defines them, are quite what we want the whole numbers to be, and I should be tempted to identify numbers with what Kneale calls quotities rather than their presentations; but I shall not go into this question now.) I welcome his refusal to regard the admitted indeterminacy, the relativity to interest or purpose, of what we take as specimens of the *same* proposition as any objection to the philosophical use of the concept of propositions ([1972*a*]). Kneale's way of defining truth, unlike that of Tarski (and, recently, Davidson) defines the ordinary concept of truth without which it is difficult to maintain clarity and accuracy in any field of thought; and he is right in saying that formalised languages are to be understood as specialised fragments of an ordinary, unrestricted, non-hierarchical language rather than as exhibiting an ideal of which ordinary language, by aiming at omnicompetence, necessarily falls short ([1972*a*], p. 243).

To sustain this line of thought, it is essential to show not only that it is proof against the well known paradoxes, but that it can resolve them in a more illuminating way than any rival view. And though I think that the main lines of Kneale's treatment of the paradoxes are correct, it seems to me to be incomplete; some details call for fuller presentation, and some further points need to be clarified and stressed.

Discussing the traditional formulation of the Liar paradox, 'What I am now saying is false', Kneale says that if we read this as 'The sentence of which my present utterance is a specimen is false', this is itself false (because *sentences* can't be false) but non-paradoxically so: we cannot go on to argue that it is therefore true. But if we read it in the paradox-generating way as 'The proposition of which my present utterance is a specimen is false', Kneale argues that no specimen of this sentence is a specimen of any proposition. Therefore there is no question of its being either true of false, so we are not forced to say that it is both. His reason is that the phrase 'What I am now saying' purports to designate a proposition (the one expressed by the whole utterance), but if it did so, it should be possible to designate it by a that-clause. But when we try to do this, we get an uncompletable formula, 'that it is false that it is false that...' ([1972*a*], p. 242).

This argument rests not only on the distinction between propositions and sentences but also on the claim that a genuine proposition should be expressible by a completed that-clause; this in turn is backed by 'the fundamental principle of semantics that for propositions and predicates alike expression is logically prior to designation' ([1971], p. 335); we must be able to express a proposition independently of designating it; no state of affairs is presentable only by talk about the presentation of it.

These comments dispose of the paradox in its traditional form. But when Kneale says ([1971], p. 322) that the 'mathematical' or 'logical' paradoxes as well as the 'semantic' ones 'all alike originate in misuses of language', we should note that the misuse is of a special sort: what it violates is not in any ordinary sense a linguistic rule, but a principle of what Kneale appropriately calls 'metaphysical common sense'; it is not merely that no state of affairs is presentable only by talk about its presentation, but further that *no state of affairs can involve its own presentation.*

This point can be illustrated by the use of a variant of the Liar paradox which the principles to which Kneale explicitly refers would not solve. Consider a sheet of paper on which are written only the following two sentences: 'Two and two make five' and 'There is no sentence written on this sheet which, standardly construed, would express a true proposition'. We are inclined to say that it must be the case either that there is, or that there is not, a sentence written on that sheet which, standardly construed, would express a true proposition. But if there is, it cannot be the first (for clearly that would express a false proposition) or the second (for that would contradict the present hypothesis); so there cannot be such a sentence. And yet if there is not, *we* must say 'There is no sentence written on this sheet which, standardly construed, would express a true proposition'; how then can we deny that the second sentence is such a sentence after all?

This more obstinate variant involves no purported designation of a proposition; it contains no reference to 'this proposition', 'what I am now saying', or the like. The only reference is to *this sheet,* and by hypothesis there undoubtedly is a sheet. Consequently the issue about which the paradox arises can be stated in a completed that-clause, 'that there is no sentence written on this sheet which, standardly construed, would express a true proposition'. The distinction between propositions and sentences is respected, and truth is applied only to the former, and there is no violation of the principle that expression is prior to designation.

What we must say here is that the apparent issue whether there is or is not a sentence written *etc.* is not after all a genuine issue because the supposed state of affairs, there being such a sentence, would involve the (negated) presentation of that state of affairs itself. Because this is not a genuine issue, the contradiction about it into which we seem to be forced is not a genuine contradiction. But to see that and why it is not a genuine issue we need to rely on metaphysical common sense, not merely on the *semantic* principle that expression is prior to designation.

In discussing Russell's paradox, Kneale concentrates not on the well known class version but on Russell's earlier formulation about the *predicate* 'not predicable of itself'. He also argues, correctly, that the crucial fault is present equally in an affirmative counterpart; we must be able to show that 'The predicate "is predicable of itself" is predicable of itself' does not express a proposition any more than 'The predicate "is not predicable of itself" is not predicable of itself' does.

To show this, Kneale gives, using a symbolic formulation, an argument which in words would rest essentially on the point that whereas, *e.g.*, 'The predicate "is human" is predicable of the predicate "is human" ' can be reduced, by the application of an obvious rule, to 'The predicate "is human" is human', an attempted reduction by the same procedure of 'The predicate "is predicable of itself" is predicable of the predicate "is predicable of itself" ' leaves us exactly where we were. He concludes from this that in the latter sentence the subject phrase 'cannot designate . . . a predicate' ([1971], p. 336).

But this conclusion, and the argument for it, are obscure. In 'The predicate "is human" is predicable of itself' the phrase 'is predicable of itself' does present a predicate; why should not *this* predicate be the one designated by the subject phrase in the problem sentence? Kneale says that this phrase 'cannot designate anything unless there is one and the same predicate expressed by the predicative phrase "is predicable of itself" whenever this occurs in predicative position in a token of a propositional phrase' ([1971], p. 335). But what justifies this ruling? And even if it is conceded, how does the above-mentioned failure of reduction prove that the phrase does not always express the same predicate?

Kneale says that 'every reflexive thesis expressible by use of "κ"—*i.e.* "is predicable of"—can be expressed also in a fashion in which its predicate can be seen clearly to be the same as its subject and therefore different from the predicate of any reflexive thesis with a different subject' and hence ' "*xκx*" cannot properly be taken as the expression of a single predicate' ([1971], p. 336). That is, because 'The predicate "is human" is predicable of itself' can be expressed by 'The predicate "is human" is human', and 'The predicate "is brown" is predicable of itself' can be expressed by 'The predicate "is brown" is brown', 'is predicable of itself' expresses *different* predicates on these two occasions. But we could argue in like manner that because 'The bus moves itself' can be expressed by 'The bus moves the bus' while 'The car moves itself' can be expressed by 'The car moves the car', 'moves itself' must express different predicates. Yet there is a clear sense in which 'moves itself' expresses the same predicate in both these sentences. We need be no more fussy about the identity of predicates than we are (cf. *above*) about the identity of propositions: the identity can be relative to and sufficient for the purpose in hand, whatever that may be. In any case, the alleged lack of a single predicate 'is predicable of itself' has no obvious connection with the above-mentioned failure of reduction, which must surely hold the key to this paradox.

The point which I think Kneale is getting at can be put in a different way. There is a predicate 'is predicable of itself': this phrase often succeeds in expressing a predicate, and what is, for some purposes, one and the same predicate every time. But this is a predicate of a special sort: a *derivative* one. What it is for a certain predicate to be predicable of itself must derive from what predicate that predicate is. It is this that fails in the problem sentence 'The predicate "is predicable of itself" is predicable of itself'. This sentence would express a proposition only if the subject phrase expressed a predicate when predicated of itself, *i.e.* of *the predicate 'is predicable of itself'*. It cannot do so, because there is here nothing appropriate for it to be derivative from. The unsuccessful reduction is the attempt to derive the required predicate from the supposed predicate expressed by the subject phrase—that is, from itself—and the obviously vicious circularity defeats this attempt.

An analogous case may further elucidate this notion of a derivative predicate. 'Is big' or 'is a big one' is a predicate, and in a clear enough sense the same predicate every time; yet it is also in an equally clear sense a variable predicate; it involves an operation on a predicate and therefore requires another predicate in the subject—*e.g.* 'This flea is a big one' or 'This elephant is a big one', meaning 'This flea is bigger than most fleas' and so on. In a similar way 'is predicable of itself' is a variable predicate, it involves an operation on a predicate, and so requires another predicate as its subject. Now if we tried to use 'is a big one' itself as the predicate in the subject—'This big one is a big one'—we should end up with no genuine predication: we should be saying that this is big for a thing that is big for a thing that is . . . *ad infinitum*. The failure of 'The predicate "is predicable of itself" is predicable of itself' is analogous to this.

Kneale seems to have been misled by a false analogy into thinking that just as there is no such thing as *the proposition that this proposition is true* so there is no such thing as *the predicate being-true-of-itself*. But the true analogy is that there is no such thing as *the predicate being-true-of-this-predicate*. There is in general a *predicate being-true-of-itself*, which, however, disappears in the special case where it would become *the predicate being-true-of-this-predicate*. (The negative counterparts can, of course, be added in every case.) This is analogous to the fact that there is in general a genuine issue whether there is a sentence on this sheet which, standardly construed, would express a true proposition, but that this issue disappears when it would become the issue whether this issue itself is settled this way or that.

We saw that a comprehensive solution of the Liar paradox required an appeal to metaphysical common sense rather than to semantics, and we now find that the same is true of the predicate version of Russell's paradox. In this case the principle required is that though a predicate can be derivative, it cannot be derivative from itself (or from its own negation).

Kneale does not show very explicitly how his solution applies to the class version of Russell's paradox. Presumably he would say that just as there is no single predicate 'is predicable of itself', so there is no single property *being a member of itself* (or, again, not being so) and consequently no class determined by such a property. But once more it would seem to be inconsistently fussy to deny that, *e.g.* the class of men and the class of trees share what is, from a defensible point of view, a common property of not being members of themselves. Again the point is rather that this property, though real enough, is a derivative one, and is real only where there is something appropriate from which to derive it. Where there is nothing to derive it from but itself (or its absence) it cannot be derived. There is no fully determinate class of all and only the classes that are (or of those that are not) members of themselves because at certain points the property of being (or not being) a member of itself becomes indeterminate. But this means that we can retain the principle that there is a determinate class for every determinate property, and hence for every determinate predicate.

Kneale notes that the heterological paradox can be expressed with an explicit distinction between predicative phrases and predicates: 'is heterological' can be defined (for predicative phrases) as 'does not satisfy the predicate it expresses' ([1971], p. 337). This shows again that it is not this and the related distinctions alone that solve the paradoxes: we need other principles as well. Again Kneale argues from the failure of a reduction to the conclusion that there is no single predicate 'is autological' or 'is heterological'. But I would prefer to say that while each of these phrases often expresses one and the same derivative predicate, they fail to express any predicate when applied to themselves or to one another because in these cases the required derivation is impossible. But again I would accept Kneale's main thesis, that this paradox, like the others, can be solved without abandoning ordinary language for a hierarchical system of languages and metalanguages. Any failures of correspondence between sets and predicative phrases (or, as Quine puts it, open sentences) are due not to any failure of the plausible metaphysical principle that there is a determinate set for every determinate property, but to the fact that certain derivative properties and predicates, which certain predicative phrases usually succeed in expressing, become indeterminate where their derivation procedures go wrong.

Kneale deals very briefly with Berry's paradox about 'the least integer not nameable in fewer than nineteen syllables'; '. . . there is no limit to the developments of notation we may introduce', so there is no largest integer which we may designate *somehow or other* in fewer than nineteen syllables, and consequently no least integer not so nameable. To solve this paradox we need no more than 'a recognition of the power of indefinite self-enrichment which belongs to all natural languages' ([1971], pp. 322–3).

This certainly disposes of Berry's paradox as stated. But Max Black gave ([1933], p. 98) what he regarded as a reformulation of Berry's paradox, which is not solved so easily, and which we should perhaps therefore recognise as a distinct paradox, and call it Black's: 'The least integer not named in this book.' Since this refers to what is not named, not to what is not nameable, the power of self-enrichment and possible developments of notation are no longer relevant; we are inclined to think that for any book there must be a largest integer actually named in that book, and therefore a least one not named there. But if the book contains some such phrase as this, there could not be (for if there were, it would both be, and not be, named in the book).

However, this paradox can be solved by the same principles as those discussed above. Suppose that the largest integer 'named' in the book in any less equivocal way is, say, 256. Then the problem phrase purports to make the naming of the next integer, 257, dependent (inversely) upon itself: if it isn't named by this phrase, then it isn't named in the book, so it is named by this phrase; but if it is named by this phrase, it is named in the book, and so it isn't named by this phrase. But no real state of affairs can be thus dependent on itself, and no real issue about states of affairs can be thus dependent on its own settlement. Consequently, there is no real issue whether 257 is named in this book or not, and the problem phrase fails to define an integer because it is indeterminate what is the largest integer actually named.

Kneale does not explicitly discuss Richard's paradox. But I would argue that what is most deeply paradoxical about it can be resolved along the same lines as Black's paradox. (I discuss this and many related points in my [1972].)

The paradoxes, then, can be solved in an illuminating way by using the distinctions that Kneale has stressed. They do not, therefore, undermine ordinary language or natural ways of thinking (contrast Quine [1966], p. 18) or the straightforward and unrestricted definition of truth which commends itself to Kneale and others (cf. Williams [1969], Mackie [1970]). In solving them we have used not only the semantic principle that expression is logically prior to designation but also the metaphysical principles that no state of affairs can either involve its own presentation or be dependent on itself, that no issue can depend on its own settlement, and that no predicate can be derivative from itself. But this appeal to metaphysical common sense only carries Kneale's own line of thought, I hope, a little further.

J. L. MACKIE
University College, Oxford

REFERENCES

Black, M. [1933]: *The Nature of Mathematics.*
Kneale, W. [1971]: 'Russell's Paradox and some Others', *British Journal for the Philosophy of Science*, **22**, pp. 321–38.
Kneale, W. [1972*a*]: 'Propositions and Truth in Natural Languages', *Mind*, **81**, pp. 225–43.
Kneale, W. [1972*b*]: 'Numbers and Numerals', *British Journal for the Philosophy of Science*, **23**, pp. 191–206.
Mackie, J. L. [1970]: 'Simple Truth', *Philosophical Quarterly*, **20**, pp. 321–33.
Mackie, J. L. [1972]: *Truth, Probability and Paradox.*
Quine, W. V. O. [1966]: *The Ways of Paradox.*
Williams, C. J. F. [1969]: 'What does "*x* is true" say about *x*?', *Analysis*, **29**, pp. 113–24.

ONTOLOGICAL VARIANCE AND SCIENTIFIC OBJECTIVITY

1 INTRODUCTION

In an earlier paper in this *Journal*[1] I showed that change of reference of the same terms in different theories is not enough to show that the two theories cannot be in conflict. Referential variance is compatible with inter-theoretical comparability. The argument in that earlier paper was an extension of the argument of Scheffler[2] that sense variance does not prevent inter-theoretical comparison. In this paper the argument is extended further. I will show how given certain changes in the ontologies of two theories my earlier argument will not work. Nevertheless, a new argument can be given to show that even certain ontological changes are compatible with theoretical comparison.

---

[1] Cf. my [1971].

[2] Cf. Scheffler [1967].

2 THE EARLIER ARGUMENT

Suppose that T_1 (combined with auxiliary hypotheses) entails

$$Ba \tag{1}$$

and theory T_2 (plus auxiliary hypotheses) entails

$$\sim Ba. \tag{2}$$

Then it would seem that T_1 and T_2 are in conflict. However, suppose we learn that the sense of 'B' has changed from T_1 to T_2 and that the referent of B has also changed from T_1 to T_2. Suppose further that we learn that every term in T_1 that is in T_2 has changed its sense and referent and every term in T_2 that is in T_1 has changed its sense and referent. Then it might well be supposed that T_1 and T_2 are incommensurable theories; they could not be in conflict.

However, this is not so as long as certain relations hold between the referents of the terms in T_1 and T_2. For example, suppose that the class of entities β is the extension of 'B' in T_1 and the class of entities γ is the extension of 'B' in T_2. Then as long as

$$(x)\,(x \in \beta \supset x \in \gamma) \tag{3}$$

T_1 and T_2 are in conflict relative to (3). Thus as long as we know that the extension of 'B' in T_1 is a subset of the extension of 'B' in T_2 the two theories can be shown to be in conflict.

3 THE EFFECT OF ONTOLOGICAL CHANGE ON THE ARGUMENT

Suppose we learn that 'a' in T_1 refers to a different sort of entity from that referred to by 'a' in T_2. ('a' is thus used ambiguously in the two theories and subscripts should be employed, *i.e.* a_1 and a_2.) Suppose further that the ontologies assumed by the two theories change radically. Suppose in particular that no statement like (3) could be formulated because the variables of the two theories range over different sorts of things. For example suppose that the entities assumed by T_1 are physical objects while the entities assumed by T_2 are processes. In this case T_1 (plus auxiliary hypotheses) would entail that a physical object named a_1 has property B while T_2 (plus auxiliary hypotheses) would entail that a process named a_2 does not have property B. In this case a statement like (3) could not be formulated since (3) assumes the values of the variables of the two theories are of the same kind. It assumes in particular that the set of entities β (the extension of 'B' in T_1) is a subset of the set of entities γ (the extension of 'B' in T_2). But the set of β physical objects cannot be a subset of the set γ of processes.

One might suppose that with this sort of change no comparison between the two theories is possible. However, this is a mistake. So long as we can find certain relations between the ontologies of the theories comparison can still be made. What is needed is some sort of ontic bridge law relating the ontologies of the two theories. For example, consider the following ontic bridge law:[1]

$$(x_1)(y_2)(Bx_1 \,.\, Rx_1y_2 \supset By_2). \tag{4}$$

---

[1] The notion of a bridge law is, of course, not new. Nagel ([1961]) Braithwaite ([1955]) and others have considered such laws. However, these authors have not, to my knowledge, considered bridge laws of the type suggested here: laws in which different variables in the law range over different types of entities. Furthermore, the use they have made of bridge laws is different from the use suggested here, namely comparison of ontologically different theories.

(The subscripts in the variables indicate that the two variables range over entities of different sorts.) Now given (4) as an auxiliary hypothesis and

$$Ra_1a_2 \tag{5}$$

as an auxiliary hypothesis we can show that T_1 and T_2 are in conflict relative to these hypotheses.

4 INCOMMENSURABILITY AGAIN?

An objection might be raised to the above argument. It might be argued that people who are committed to T_2, for example, will not accept an ontic bridge law like (4) or a hypothesis like (5). The advocates of T_2, it might be said, do not believe that physical objects exist. Hence it might be argued, the problem of the incommensurability of the two theories remains.

It might well be true that the advocates of T_2 would not believe that physical objects exist. But this is not relevant to the point at issue. If they believed this, then perhaps from the point of view of *their* ontology no comparison between T_1 and T_2 could be made. Since if they accept only processes as real, then it might be supposed that they could not believe that some physical object a_1 was in a relation R to a process a_2. That is, they could not accept (5). But this would not mean that *no* comparison was possible or that the ontology of the advocates of T_2 was correct. *We* might believe that physical objects exist and thus believe that a_1 stands in relation R to a_2.

Furthermore, there is no *a priori* reason to suppose that the advocates of T_1 would not accept an ontic bridge hypothesis like (4) or a hypothesis like (5). There might be good reason to induce the advocates of T_2 to enlarge their ontology to include physical objects and thus to consider ontic bridge hypotheses like (4) and hypotheses like (5).

Notice also that to speak of 'the acceptance of hypotheses' like (4) and (5) is ambiguous.[1] In one sense 'the acceptance of a hypothesis' means 'believing the hypothesis is true' but in a weaker sense it simply means 'consider the hypothesis as possibly a fruitful research tool.' In this second sense of 'acceptance' no belief in the truth or falsity of the hypothesis is presumed. So even if people who believe that T_2 is true could not bring themselves to believe that there are such things as physical objects they might still accept (4) and (5) in the weaker sense of accept. They might suppose that (4) and (5) are useful research hypotheses and consequently suppose that it is useful to work with the hypothesis that there are such things as physical objects even though they did not believe that there are physical objects. So construed the advocates of T_2 could accept (4) and (5) and make a comparison of some kind between T_1 and T_2.

One interesting application of this idea is in the following case: Suppose that T_2 entails that 'B' in T_1 has extension zero and/or 'R' has extension zero.[2] Advocates of T_2 clearly could not accept Ba_1 in the strong sense of 'accept'. Nevertheless, they might still accept Ba_1 in the weak sense; they might accept Ba_1 as a useful research hypothesis. Advocates of T_2 could even accept (4) in the strong sense of 'accept' despite the fact that 'R' had no extension in their theory. In general the falsehood of the antecedent of a law does not falsify the law. However, even if advocates of T_2 thought (4) was false they could still accept (4) in the weak sense of 'accept'.

---

[1] Cf. Scheffler [1967], p. 86.

[2] Cf. Giedymin [1970], p. 266, footnote 1.

Other possibilities might arise which would make comparison more difficult. But each case would have to be evaluated on its own merits. No general argument seems possible to show that such difficulties need always arise or that if they arise, they could not be handled. Suppose for example that theory T_2 entailed that 'B' in T_1 was meaningless.[1] In this case advocates of T_2 could not even consider Ba_1 as a useful research hypothesis. However, this does not mean that *we* could not have a criterion of significance that would allow Ba_1 to be a meaningful hypothesis and hence that we could not consider T_1 and T_2 as comparable *via* ontic bridge laws like (4).

Another possible problem is suggested by some examples of Feyerabend.[2] The bridge law (4) could be incompatible with either T_1 or T_2. (The cases we have considered so far have been those in which (4) is incompatible with the conjunction of T_1 and T_2.) It might be argued that in this case (4) could not be consistently accepted by advocates of either T_1 or T_2. However, although advocates of T_1 or T_2 might not be able consistently to accept (4) in the strong sense of 'accept', they might still accept (4) with perfect consistency in the weak sense of 'accept'; thus comparison could still be made. Furthermore, there is some problem in seeing how (4) could be logically incompatible with T_2. For (4) (in extensional terms at least) could only be false in terms of T_2 if T_2 entailed that the antecedent of (4) was true and the consequence of (4) was false. How both of these conditions could be met is difficult to see given the ontology of T_2.

5 ARE ANY TWO THEORIES COMPARABLE?

One possible objection may be raised against the position suggested here. It might be objected that on this view any two theories are comparable relative to some bridge principle. Indeed, even apparently incommensurable hypotheses such as 'All planetary orbits are elliptical' and 'All circumferences of pennies are circular' become comparable given suitable bridge principles. For it is a trivial logical truth that given any two consistent sentences P and Q one can find another consistent sentence R such that when R is conjoined to P and Q there is an inconsistent conjunction.[3]

But what is not possible in all cases is to find an empirically true or well supported sentence R that when combined with P and Q yields an inconsistent conjunction. The ontic bridge principles referred to in this paper are special kinds of empirical laws. These principles must have all the various properties essential to laws: universality, truth, counterfactual force and testability. Such statements are not so easily found and in many cases it may not be possible to find them at all. Whether such statements can be found is not determined *a priori* but will depend on empirical investigation. Thus it is not necessarily true that on the view presented here any two theories are comparable. It may turn out that very few are. Nevertheless, it is possible that such laws could be found. Thus comparability is possible despite ontological change.

MICHAEL MARTIN
Boston University

---

[1] Cf. Giedymin [1971], p. 47.

[2] Cf. Feyerabend [1962].

[3] This point was called to my attention by Norman Swartz and Raymond Jennings.

REFERENCES

BRAITHWAITE, R. [1955]: *Scientific Explanation.*

FEYERABEND, P. [1962]: 'Explanation, Reduction and Empiricism', in H. Feigl and G. Maxwell (*eds.*): *Minnesota Studies in the Philosophy of Science*, **3**, pp. 28–97.

GIEDYMIN, J. [1970]: 'The Paradox of Meaning Variance', *British Journal for the Philosophy of Science*, **21**, pp. 257–68.

GIEDYMIN, J. [1971]: 'Consolations for the Irrationalist?', *British Journal for the Philosophy of Science*, **22**, pp. 39–53.

MARTIN, M. [1971]: 'Referential Variance and Scientific Objectivity', *British Journal for the Philosophy of Science*, **22**, pp. 17–26.

NAGEL, E. [1961]: *The Structure of Science.*

SCHEFFLER, I. [1967]: *Science and Subjectivity.*

Brit. J. Phil. Sci. 23 (1972), 257–290 *Printed in Great Britain*

Review Articles

HEGEL'S PHILOSOPHY OF NATURE[1]

The influence of Hegel's philosophy has been immense, first and foremost, of course, on the development of political theory; secondarily, on the philosophy of history, of art, and on the development of general philosophy itself; and on figures like F. H. Bradley, McTaggart, Bosanquet, Joachim, Collingwood, Croce, Royce. But just to mention these names shows how the Hegelian influence has waned in the philosophical market-places of the last fifty years: the language and problems of 'normal philosophy' (using this expression on analogy with Thomas Kuhn's 'normal science') have not been Hegel's, although some valiant attempts have been made at integration with contemporary modes of speech, notably in the writings of J. N. Findlay in England, and Martin Heidegger on the Continent. More recently, there has been a revival of interest in Germany (*e.g.* Gadamer, Habermas) which bids well to transcend mere attempts at rehabilitation within the framework of the language of established academe, and which searches for novel integration of Hegelian approaches with a widened area of the philosophy of the sciences (natural and social) as well.

In all this, Hegel's *Philosophy of Nature* has, if anything, suffered even more neglect. Composed over a span of some twenty-five years, its main outlines were conceived already during Hegel's Jena period of 1805–6. An intensely personal vision, eventually formulated as the dialectical edifice of the *Encyclopedia*, its categorial network was superimposed upon, and was intended to articulate, a body of science, itself constantly in flux, containing much that was problematic, and in the process of undergoing considerable change. No wonder that the scientists of the time, not to mention those of subsequent years, were appalled by what appeared to them a mixture of ignorance of scientific advance and peremptory dismissals of time-honoured paradigmatic approaches to areas that seemingly had achieved a venerable degree of stability, *e.g.* Newtonian physics. In this area, Hegel's attempt, even though not actually an attack on scientific achievement, but only on the scientific paradigm in which this achievement was expressed, found few sympathetic ears, least of all from a scientific public who had few clear ideas on the distinction between science and philosophy of science, let alone any understanding of Hegel's attempts to re-articulate the principles of such a philosophy, particularly when couched in his congealed and highly personal language.

As for the non-scientific reader (philosopher and layman alike), Hegel's philosophy of nature would remain mostly incomprehensible, though for a reason opposite to that of the scientist: the very virtues of Hegel's philosophy, its being soaked in the details of the science of the period between 1770 and 1830 would make it inaccessible to the philosophical reader. Hence there were few intelligent responses (cf. Petry's survey of the contemporary literature, i, pp. 115–22), whereas the charges of ignorant dilettantism hurled at Hegel both during his lifetime, but more especially in the sequel, need to be measured against the wide range of scientific sources of which Hegel had actually availed himself in composing the *Philosophy of Nature*: the list of works quoted—as collated by Petry—amounts to thirteen pages of text (i, pp. 128–41). Unfortunately, since these sources soon lost their relevance, not to say their meaning, by becoming out-of-date, one could hope to find one's way round Hegel's scientific allusions only with the help of a trained historian of science. This is the task which Petry has undertaken in his three-volume edition of the *Philosophy of Nature*, with a patience and breadth of erudition that is truly astonishing. Half of each volume consists of notes, which in addition to short paragraphs, occasionally attempting to shed some light on tricky Hegelian formulations, supplement Hegel's own stated authorities by further reference both to contemporary and later bibliographical scientific sources, as tools for reconstructing the intellectual climate within which (and against which) Hegel was operating. Moreover, the typesetting of the translation distinguishes between the parts of the text belonging to different periods of composition; the original as we now know it was published only in 1842 by Michelet, and included Hegel's main paragraphs together with his 'Remarks' as well as a vast amount of lecture-material culled from various students' lecture-notes, reproduced as 'additions' (*Zusätze*).

---

[1] Review of G. W. F. Hegel [1842]: *Philosophy of Nature* (edited, translated and with an introduction and explanatory notes by M. J. Petry), London:George All en and Unwin; New York: Humanities Press, 1970, £18.00, 3 vols., Pp. 39 2+469+422, and G. W. F. Hegel [1847]: *Philosophy of Nature* (translated by A. V. Miller with a foreword by J. N. Findlay), Oxford: Clarendon Press, 1970. £3·75, Pp. xxxi+450.

The need for such technical aids becomes evident as soon as we study even the simplest of Hegel's critical escapades. Consider, for example, Hegel's critique of Newtonian gravitational theory as presented during his life-time, a subject that will help us also to enter at once into some of the problems involved in seeking a more appropriate understanding of Hegel's philosophical intentions than has hitherto been offered to us. As is well known, Hegel rejects Newton's method of providing an *explanation* of Kepler's laws of planetary motion in terms of an 'independent' force of gravitation, still less the more popularising explanations involving the notion of spurious centrifugal forces. (Cf. i, p. 354.) Now before one can begin to evaluate Hegel's response, it is necessary to be aware of the vast background of scientific and mathematical literature available to and made use of by Hegel at the time of his composition of the *Philosophy of Nature*. This involved, among other extant theories of celestial mechanics (i, p. 346), the question of the logical rigour of the Newtonian calculus (i, p. 349), Newton's use of the principle of composition of forces (i, p. 350,) and a whole host of supplementary considerations, all indicated with admirable patience in the work under review. (Cf. pp. 343–5.) What the presence of such a literature indicates is the extreme complexity and fluidity of the scientific situation in a subject which to us may well seem to be settled, and was often believed to be similarly settled during Hegel's time.

One may go further. One may object (with our knowledge of hindsight) that many of Hegel's objections were intellectual quibbles and often irrelevant. The situation, however, looks quite otherwise if we treat it as a supremely interesting attempt to provide an alternative *paradigm*, which Hegel is here opposing to that of Newton. (After all, our own century has seen the overthrow of the theoretical edifice on which this was based.) Through Hegelian spectacles, such objections would not then seem mere 'quibbles' but serious indicators of a fundamentally 'wrong' way of looking at 'the phenomena'. Moreover, what is of the utmost importance to realise in the present example is that the paradigm which Hegel is proposing is itself rooted in a substructure that may be called Hegel's 'protophysics', where he seeks to locate the basic scientific concepts in an underlying series of conceptual explications (Hegel's *Begriffsbestimmungen*) believed by him to emerge naturally from the categorial and dialectical approach as already elucidated in the first volume of the *Encyclopedia*, the *Logic.*

Hegel did not oppose Newtonian science lightly: 'Propositions [*sc.* Hegel's] not in accordance with accepted opinion look like mere assertions', he writes, 'and when they contradict sober authorities, seem to be presumptuous, which is even worse' (i, p. 265); nor was it undertaken as a whim, since his opposition goes right back to his doctoral thesis of 1801, 'On the orbits of the planets'. Basically, Hegel's approach is not unlike that which came later to be associated with the work of Heinrich Hertz: the *fundamental* concepts to be employed in a *systematic* presentation of mechanical phenomena should be limited to space, time, matter and motion—but not include force. Like Pierre Duhem later, he moreover argues that a theory of dynamics should not be regarded as providing 'explanations', in the special sense of invoking 'hidden' entities like forces, waves and atoms, but that our theory-types for preference ought to be of (what Ernst Mach later called) the 'phenomenological type', in a sense purely 'descriptive', employing a minimum number of primitive concepts, gleaned from the realm of the 'observable'. Hegel here contrasts the presentations of Lagrange and some of the writings of his minor disciples, *e.g.* L.-B. Francoeur, with those of the Newtonians. (Cf. i, pp. 336–7, 254–5; and for Francoeur, i, pp. 332, 351.) Formally, the *basic* starting-point in these accounts are certain empirical laws, *e.g.* Galileo's law of free-fall, Kepler's laws of planetary motion. In Lagrange's 'analytical method' for instance, as Hegel remarks (i, pp. 337–8), we begin with the exposition of certain mathematical functions, and then simply seek to discover which of these describes a given physical state of affairs. The rest of the work then consists in giving a logically consistent systemic account of the interconnections and implications of this 'foundation'. True, the Lagrangian presentation may be shown to be equivalent to that of the Newtonian, but it is less misleading; for one thing, it makes force a 'derivative entity' (*ibid.*); for another, it tempts us less to give misleading *physical interpretations* of the formal calculus; the Newtonian form has been responsible—according to Hegel—for a flood of 'monstrous metaphysics' (i, p. 265).

The 'metaphysics' involved here, was of course that of the false 'reification' of gravitational force, as an 'independent' quasi-substantial entity, which fitted very badly into Hegel's protophysics. Hegel here ties up with the traditional criticism of Newton's concept of gravitational attraction, putatively introduced purely as an 'external' and *ad hoc* invention (i, p. 238), whereas if one wants to employ the concept of force one must needs—as with any basic concept of physics—show it to be embedded in the fundamental structure of experience. Hegel here makes contact with the results of the Kantian attempt to provide 'foundations' for Newtonian dynamics; an attempt which—Hegel proclaims—was tantamount to the revival of the very 'concept of a philosophy of nature' (i, p. 241). The trouble with Kant, according to Hegel, was that he had not gone far enough. Kant had provided an elucidation of our *empirical* concept of matter, according to which the power to attract and repel 'at-a-distance' is part of the very 'essence of matter'; where the analysis of matter thus provided had been shown to be complete and systematic in accordance with the Kantian categorial scheme. (This was Kant's so-called 'metaphysical foundation', *i.e.* his elucidation of the *possibility* of the Newtonian science of mechanics.) Unfortunately, Hegel complains—though for the full discussion we have to go to his *Science of Logic*—it has not been shown with sufficient clarity how one must explicate the concept of matter in a *logical* (and not just empirical) sense, without introducing determinations which involve a recourse to the familiar notions of repulsion and attraction. One cannot just *posit* that matter is impenetrable and offers inertial resistance; more generally, one cannot just posit matter in space and time, 'indifferently', instead of showing that in the evolutionary progress of the categorial exposition, at each stage the tension of opposites provides the motive power for advance to the next 'level', thus naturally leading to 'matter' and its 'ideal moments', *i.e.* attraction and repulsion.

Unfortunately, the true significance of this exposition (i, pp. 241, 243) will escape the reader unless he realises that there lies at the back of all this the parallel explication of the categorial concept of 'being-for-self', given in the *Logic* (cf. *e.g. Encyclopedia*, vol. i, *Logic*, §§96–7). Hegel there purports to show that 'being-for-self' must be construed as a 'one' and as a 'many', and that the tensions involved in maintaining a 'many-one' relationship must be expressed ('metaphorically speaking', Hegel adds) through the concepts of attraction and repulsion. Contrary to Kant, force thus is an aspect of matter in virtue of its *logical* explication as 'existent being-for-self' (i, p. 240). (This is one of many instances, where Petry's commentary, limiting itself almost wholly to scientific sources, is quite insufficient to provide a grasp of what Hegel *means* by 'philosophy of nature', and even the very full 100-page introduction to the work in volume i, whilst attempting to sketch in some of the general background to Hegel's thought, is often quite insufficient to help us grasp the structure of the argument.)

From here on, the peculiar 'paradigm' of Hegel's theory of matter and gravity unfolds in easy stages. For whilst matter is 'inseparably both' attraction and repulsion, this gives us only the 'extrinsicality' (i, p. 241; Miller, p. 44, translates 'asunderness': *Auseinandersein*), *i.e.* matter *qua partes extra partes*, so we again require the dialectical opposite of this, as its 'negative unity', the 'ideal singularity' of matter, which for Hegel is gravity, as centre of gravity (*i.e.* each continuous though finitely extended material body has its centre of gravity). And the same, of course, goes for a plurality of such bodies: they, too, in virtue of the preceding explication, will have their ideal centre of gravity, *e.g.* the centre of gravity of the solar system.

This exposition has, however, a curious consequence, albeit the one conveniently desired by Hegel to accord with his preference for phenomenological or purely 'descriptive' theory-types. Evidently, not only must the two forces always be regarded in pairs (so attraction alone can never become an independent real force), but these forces are, so to speak, locked within the bosom of matter; thus they will exhaust their 'effect' by providing for each material system its centre of gravity. Gravity thus becomes a determinate characteristic of a material system as such; in consequence, and thus differing from Kant, 'it is essential to distinguish gravity from mere attraction' (gravitational attraction); matter is—as in Aristotelian physics!—'essentially heavy' (i, p. 242, has: 'the primary essence of matter is that it has weight', but Miller translates correctly: 'matter itself is essentially heavy', thus preserving better Hegel's Greek intentions), a property which it thus—*pace* space-travellers!—can never lose. Gravity is understood as the tendency for a plurality of bodies to strive towards their centre of gravity. And this manifests itself in their relative motion towards their common centre; indeed, as Hegel remarks, 'motion as such only has significance and existence where there is a system of several bodies, which are variously determined and so stand in a certain relationship to one another' (p. 261), an important insight anticipating much later ideas on the subject.

The whole account has thus an Aristotelian flavour; it is, as it were, determined kinematically, rather than kinetically, though in the short space of this review it is not possible to indicate its subtlety and complexity. At any rate, gravitational attraction as an active agency is proscribed, even though, as Hegel notes, 'one can say that matter is attracted to the centre' (i, p. 242), but this a *façon de parler* for the phenomenological account just given. Hegel did not of course claim that his protophysical account could yield *deductively* the kind of analytical theories of which he thought Lagrange's to be the most powerful instance. For science, such an account as Hegel's—notional determination, understood as the placing of a given concept within the dynamics of the general categorial framework—could at best be only direction-giving.

> It is the impotence of nature [he writes], that it preserves the determinations of the Notion only abstractly, and leaves their detailed specification to external determination. (i, p. 215; I have again used Miller's translation, p. 23, as being more true to the flavour of the original.)

At most, the scientific edifice will exhibit 'traces of the determination by the notion', and for this reason there are 'limits set to philosophy', and it would be 'quite improper to expect the Notion to comprehend—or as it is said, to construe or deduce—these contingent products of nature' (*ibid.*).

I have discussed this example in relative detail because only thus can one get an idea of the relation between Hegel's Logic, the Philosophy of Nature and theoretical science. Philosophy of Nature is meant to provide a 'grounding' for basic scientific concepts, assessing their logical status, in the light of the general categorial approach. In this it will for the most part seek to determine no more than paradigmatic approaches for science, as in the example just given, although the degree or depth of the criticism will vary from case to case.

The Petry Introduction here usefully sketches the background to Hegel's general approach. In the section on 'levels, hierarchies and spheres' an account is provided of the structure of the *Encyclopedia*, and the impact of this on Hegel's conception of the *qualitative* difference that he believed to exist between different spheres of science, is broadly indicated. (Hegel was clearly an extreme 'anti-reductionist'.) Petry thinks that Hegel was quite open to the suggestion that the 'subject matter [of science] should be regarded as being open to constant revision in the light of the changing state of knowledge', but that it is just as essential that 'this subject matter should be structuralised with reference to the principles through which it becomes most fully intelligible', the principles being regarded 'as absolute and changeless' (i, p. 33). In a similar vein, Petry later notes the contrast between the permanent general structure and—as he somewhat doubtfully claims—the 'perpetually changing and infinitely revisable' parts of Hegel's *Philosophy of Nature* (i, p. 90). Unfortunately, the passage to which Petry alludes in order to back up this claim, does not, as he seems to suggest, refer to the 'general structure' in the sense of the 'metaphysics' of nature, but deals with the nature of scientific thinking, in the sense of scientific theorising. Hegel says: '*Jemehr des Denkens in der Vorstellung wird, desto mehr verschwindet von der . . . Einzelnheit und Unmittelbarheit der Dinge*': *i.e.* 'the more of thought is converted into representation, the more there disappears of the singularity and immediacy of things', and here '*Vorstellung*' or 'representation' is Hegel's technical term for 'abstract' scientific laws and theories, akin to his notion of 'reflection' as scientific thinking (cf. i, p. 110). Petry translates 'The more thought predominates in ordinary perceptiveness . . . *etc.*' (i, p. 198), which betrays an astonishing lack of sensitivity towards Hegel's meaning; Miller, p. 7, has: 'The more thought enters into our representation of things, . . . *etc.*'—not quite what Hegel says, but at least closer to his intentions.

The rigidity of this distinction, the degree of 'absoluteness' attaching to the protophysical substructure, the extent to which it should or can dictate the theoretical (scientific) superstructure, is of course problematic, and Petry's commentary does little to help us solve its problem. The minimum that one could perhaps claim is that the substructure may in times of revolutionary change provide intellectual anchors and points of view, whereby new concepts whose significance is not yet too clear—though their scientific fruitfulness may be temporally acknowledged—may come to be given some kind of intelligibility. Evidently, protophysics will here have a tendency to act in a conservative fashion though in its direction-giving function it may also occasionally be the harbinger of later revolutions. Anyone who reads Hegel's pronouncements on space, time, matter and motion in an open-minded way cannot but be struck by the glimpses they afford of subsequent twentieth-century approaches. Not that Hegel here, for instance, *anticipated* relativity theory. Nevertheless, his philosophy provides signs through which the general attitude of relativists can be shown to connect intimately with possible alternative conceptual explications, and such explications have not unfrequently been thought to form an important task for science; cf. Whitehead's conception of 'pan-physics' in his *Principle of Relativity*, where the concepts of relativity, such as event, space-time, particle, field, *etc.*, are constructed by being related to a broad metaphysical background of the structure of what he calls 'the relatedness of nature'.

Petry himself likens Hegel's notions of levels, hierarchies and structures to ideas found in certain contemporary writings, including names like Kapp, Kemeny, Putnam, Woodger, Bunge, Braithwaite, Harré and Tisza; with a notion like Harré's 'general conceptual system' (i, p. 38) being suggested as a kind of third-cousin to the more broadly-based approach of Hegel. Petry explores briefly the various areas of Hegel's critique of certain scientific theories of his time, in the fields of mechanics, dynamics, optics, biology, and especially emphasises the important lessons to be derived from Hegel's attempt to set the various branches of science against one another, drawing attention to the need both to develop these independently and to allow subsequent cross-fertilisation at a peculiar stage of their development. (Cf. i, pp. 58–60.) (Although one cannot but be sanguine about any awareness on the part of Hegel concerning the need for the second stage, cross-fertilisation. Berzelius's electro-chemical theory is after all criticised by him precisely for its *mixing* of electrical and chemical conceptions! Cf. ii, pp. 200–13).

On the other hand, Hegel's opposition to Daltonian atomic theory (the 'worst form of atomistic metaphysics', ii, p. 213) will be more easily appreciated as yet another attempt to interpret phenomena (this time chemical phenomena) through the spectacles of a purely 'descriptive' theory, Hegel, of course, siding with the notion of 'chemical equivalents' (on the lines of Humphry Davy for instance) rather than relative atomic weights, although for a more extensive discussion of this one must again refer to Hegel's Logic, such as the 'Remark' in the *Science of Logic* about 'elective affinity' appended to the section on the category of Real Measure.

Similarly, Hegel's celebrated objections to orthodox theories of the refraction and dispersion of light must be viewed as dominated both by the position accorded to 'light' in the scheme of the categories, and by Hegel's corresponding preference for 'descriptive' theories—though he managed no more than a qualitative sketch. (He probably would have welcomed Fresnel's subsequent 'mathematical' treatment.) Here, again, to grasp Hegel's discussion of refraction (ii, pp. 125–33), Petry's notes are a minimum essential (ii, pp. 340–8), although his explanation in the Introduction, of the general philosophy lying behind this topic, leaves something to be desired (cf. i, p. 55, also ii, p. 348): Hegel's 'descriptive' approach can hardly have 'kept the way open for' the demonstrations of the undulatory theory of light (i, p. 55; cf. ii, p. 348), considering that he holds that 'the wave-theory . . ., *e.g.* Euler's ether and vibration of sound, are material representations which are quite useless for the understanding of light' (ii, p. 20). Hegel's 'account' of light in the categorial scheme is simply not consistent with physicalisation, and he realises of course the difficulties his own theory involves, with its non-Newtonian conception of 'specific gravity', in the end admitting that 'we have to be content with keeping to and establishing a general point of view' (ii, p. 133). That point of view was of course admirably suited to harmonise with Hegel's explicitly stated intention of excluding consideration of the *material* nature of light (cf. 'Fermat's Principle of Least Time'), and we must imagine Hegel making propaganda for this kind of approach. (Petry wrongly cites Fermat as one of the defenders of 'mechanical explanations' known in Hegel's time to be 'inadequate', ii, 347.)

It is difficult here to retain one's perspective. Admittedly Hegel was no theoretical scientist, although he *understood* the writings of other scientists, and was only too able to utilise for his own purposes their shortcomings, and their criticisms of the shortcomings of each others' theories. On the other hand, his approach is interesting in so far as it points to examples of *different* 'thinking-caps', although his continual Aristotelian classifications (Duhem was later to return in a similar way to an Aristotelian foundation of science) tend to be confusing to the modern reader, not to say otiose.

At any rate, we must regard the whole advance of the 'idea' in its 'external' garb as 'nature', from 'mechanics', *via* 'physics' to 'organics', as being articulated (though only approximately) by the structure of the underlying 'logic'. And the 'mediation of opposites', for instance, will be a locution applied throughout, whether we are concerned with such opposites as attraction and repulsion, at one extreme, or with the sexual union between male and female, at the other, where

> the *Idea of nature* is actual in the male and female couple; up till now their *identity* and *their being-for-self* merely had being for us in our *reflection*, but they are now experienced by the sexes themselves in their *infinite reflection* into each other. This feeling of *universality* is the supreme moment of the animal's capabilities, but within it, its *concrete universality* never occurs for it as a theoretical object of intuition. If it did, it would be *thought or consciousness*, in which alone the genus attains free existence (iii, p. 173; my italics).

This passage may serve as a picturesque example—and it is here *only* a picture—of this method of articulating the processes of nature *in terms of* the abstract concepts of logic, infused into the more concrete and empirical concepts of science, though for the deeper significance of this method it is best to return to contexts like those I discussed at the start, in connection with Hegel's approach to force and theoretical astronomy.

Petry's Introduction points out that a full understanding of this volume of the *Encyclopedia* requires reference to its coping-stone, the third volume, on *Spirit*, and he gives a useful account of the connections between the three tiers of Hegel's *Encyclopedia* (i, pp. 91–3), followed by a discussion of the general range of conceptions like Idea, Notion and the dialectic (i, pp. 93–7). This account may be usefully supplemented by Findlay's more scintillating introduction to the Miller translation, written with the usual elegance and inimitable mastery of Hegel's thought that one has come to expect from this author. Petry duly notes what he calls the Hegelian 'counter-Copernican revolution', its placing of the sentient sphere above the inanimate, of the solar system above that of the fixed stars, *etc.* (i, pp. 105–8), in line with Hegel's general thesis that Spirit may well be regarded as capable of changing the categories of science, and thereby 'assimilating nature' through its 'idealisations', so that in the end 'nature is posited by spirit' (cf. *Zusatz* to §381, *Philosophy of Spirit*). (A modern vulgar rendering of this would be that scientific research councils, through their funding operations, create the intellectual and material basis for the realisation of research programmes that yield paradigms productive of our pictures of nature.)

Hegel has some perceptive things to say on paradigmatic thinking in science, and we come back in Petry's closing pages to the problem of the relation between the notional determinations of his dialectical logic and the theories of science (i, pp. 109–12). In the end, we may say that all that Hegel claimed was the right, on the side of 'philosophy', to question general paradigms, and his remarks have a poignant reading in the context of the situation in which philosophy of science finds itself in our own time:

> [Scientific professionals] often want to form a caste, and to be in exclusive possession of their science, and permit no one else to form a judgement.... Within such a caste, certain fundamental conceptions are developed which mould everyone's thinking. If one does not follow their language, he is supposed not to understand the subject, as though only the members of the guild had the appropriate understanding. Now, they are right: The [scientific] *Understanding* of that subject, *this* category one lacks—this [type of] metaphysics, according to which the subject is supposed to be viewed. Philosophers especially are here rejected; [but for this reason] it is precisely

> they who have to attack those categories. (Cf. Petry, i, p. 109; ii, p. 154; Miller, pp. 211–12; the context is the discussion of Newtonian approaches to the theory of colours.)

I have translated this passage afresh, since neither of the translations under review seem to bring out with sufficient pithiness Hegel's play on the word 'Understanding' (although Miller at least spells the word with a capital U). 'Understanding' denotes the scientific approach (and here, in particular, Newton's theorising on dispersion, employing the theoretical conceptions of coloured rays), unmediated, involving 'mere reflection', and clashing with the conceptual determinations suggested by the dialectical logic. So Hegel makes the general claim that competing paradigm-defenders do not speak each others' language, and the particular claim that any new paradigm, competitive with the Newtonian, ought to take into considerations the phenomenological approach implied by his logical articulation.

In general, although neither translation can of course hope to break into the closed circle of the German original, I have found Miller's translation frequently to give a more faithful rendering than Petry's; and in any case, it is best to regard translations of Hegel as a mere guide to the original, and not the last word. But both translations on occasion unfortunately make a difficult text still harder. Consider a passage from the *Zusatz* to §264: '*Die Dauer ist Ruhe in der Beziehung, dass sie, als Begriff, ihrer Realisirung, der Bewegung engegengesetzt wird.*' Now literally, this means something like the following: 'Duration is rest, in this respect, that it, *qua* Notion, is being opposed to its realisation, to motion.' Petry has: 'Duration is rest in that as the Notion of its realization, it is opposed to motion.' Evidently, Petry has forgotten about the comma between '*Begriff*' and '*ihrer*', thus producing the nonsense 'Notion of its realization'. Miller translates: 'Duration is rest, in the relationship in which the former, as Notion, is opposed to its realization, to motion.' Is Miller being oversubtle, in translating '*in der Beziehung*' as 'in the relationship in which'? True, '*Beziehung*' *can* mean relation, and there is a faint possibility that Hegel uses the term to echo the fact that the Notion of duration, is *qua* Notion, being 'related' to its opposite, rest. But I do not think that any person using the colloquial locution '*in der Beziehung*' in the present context would mean this. Quite generally, it will be agreed that neither translation exactly facilitates an understanding of the flow of Hegel's thought.

There are many such cases; I will limit myself to mention of one more, since it involves also a matter of principle. In the famous §270, Hegel, *after* some critical remarks on the Newtonian account of planetary motion, *sets out* to give a *notional* account of the phenomena described by Kepler's laws, in contradistinction to Newton's 'proof' of Kepler's laws. He introduces this as follows:

> *Wenn nun hiernach einige Grundzüge angegeben werden, wie die Hauptbestimmungen der freien Bewegung mit dem Begriffe zusammenhängen: so kann diess für seine Begründung nicht ausführlicher entwickelt . . . werden.*

Literally translated:

> If now, in what follows, some general indications are given of the way in which the main determinations of free motion connect with the Notion, then this cannot, as regards its grounds, be developed in any greater detail.

Miller translates: 'If now *in this way* [my italics] some main features are indicated . . . *etc.*' Now 'in this way' suggests that Hegel is talking about what has preceded, and the reader will hence not understand that this important paragraph relates to what follows.

Petry's translation makes the same mistake, only more explicitly: 'The connection between the main determinations of free motion and the Notion *has only been* outlined [my italics], and as it is not possible here to develop it any further . . . etc.' The structure of the argument is clearly destroyed.

How could this have happened? Perhaps because Hegel says: '. . . this cannot be developed in any greater detail', suggesting that what has *preceded* cannot be developed further. Is this only a pedantic quibble about accuracy of translation? Not at all! For the whole point of Hegel's thought lies in what *follows*: to give a sketch of what a *notional account* of scientific connections ought to look like, as distinct from his critical assessment of the Newtonian formulations and his more positive assent to Lagrangian presentations. Hegel's notional account is not *meant* to be in competition with either; but that idea is concealed by the above-mentioned mistranslation, possibly because our translators here lack a sufficiently clear grasp of the subject-matter.

The above misunderstanding is much like the common view that Kant sought to supply an *a priori* demonstration of the law of gravitation, because he suggested (what he calls) 'a possible construction', *i.e.* a mathematical model of gravitational action simulating an inverse-square effect. Kant's notion of 'possible construction' was, as has been indicated, the predecessor of Hegel's 'notional determination', and one would therefore not expect Kant to want to compete with those scientific processes that lead to the inductive foundation of laws. And we shall not therefore be surprised when we find Kant indeed say that the law of gravitation, as a physical truth, can only be accepted as a conclusion from the data of experience. The contrary legend is once again due to an insufficient distinction between Kantian 'physics' and 'protophysics'. The charm of Hegel's philosophy of nature lies in providing us with a vast fossil specimen of protophysics, perhaps indicating in firmer outline the nature of this concept in proportion to its severance from the scientific orthodoxies of the time.

It is perhaps an indication that this significance of Hegel's work is beginning to dawn on philosophers of science that two translations should appear in one year, as an expression of the fact that we may be about to enter into a new era of the philosophy of science.

GERD BUCHDAHL
Department of History and Philosophy of Science,
Cambridge University

SCIENCE, HISTORY AND METHODOLOGY

What is the relation (or lack of relation) between philosophy of science on the one hand and history of science on the other hand? Reflection on this issue is stimulated, in me at any rate, by the volume *Criticism and the Growth of Knowledge*[1] and by some of the contributions (especially 'History of Science and its Rational Reconstructions' by Imre Lakatos) to the eighth volume of *Boston Studies in the Philosophy of Science*.[1] (I shall abbreviate the titles of these volumes by '*CGK*' and '*BS*' respectively.) With apologies to the other contributors to the volumes, I shall concentrate on those papers which seem to me to deal most directly with this particular problem.

---

[1] Imre Lakatos and Alan Musgrave (*eds.*) [1970]: *Criticism and the Growth of Knowledge. (Proceedings of the International Colloquium in the Philosophy of Science, London,* 1965, 4.) London: Cambridge University Press. Second impression 1972. £1.40 pp. viii+282.

In many universities there are departments of History and Philosophy of Science, but I have myself for some years been sceptical of the value of so juxtaposing history and philosophy. Many of the activities of historians of science have been without philosophical interest, just as many of the activities of philosophers of science have been without historical interest. Moreover I cannot help feeling that the lure of history has sometimes directed first rate philosophers of science from the paths of philosophy. As against this, some philosophers of science have used the history of science to illustrate and to support interesting and important philosophical theses. Similarly Kuhn, though primarily a historian, uses the history of science to raise important philosophical questions. So it *seems* that history of science *can* be philosophically relevant. But is this appearance perhaps illusory? Ever since I first read Popper's *Poverty of Historicism* I have been sceptical about the possibility of learning any predictive lessons from history. Many philosophers of science, however, do seem to think that we can learn from the history of science what is the right way to go about science now and in the future. Is this not historicism? It is therefore noteworthy that Imre Lakatos, though he is much concerned with the history of science and with the historiography of science, explicitly repudiates historicism in this area. See *BS*, p. 179, where he refers to Popper's *Poverty of Historicism* in just this connection. Let us investigate whether he manages to be quite successful in avoiding historicism in his own methodological theorising.

There are two features of Lakatos's philosophy which lie behind his rejection of historicism. The first of these is his distinction between internal history and external history. (Kuhn remarks, in *BS*, pp. 140–1, that Lakatos uses these terms rather differently from the way in which historians of science have customarily used them.) Internal history is history reconstructed as if the scientist had always proceeded in a rational manner, according to the principles of some methodology. Since different sorts of philosophers have different ideas as to what constitutes rationality (what is the correct methodology) it follows that their internal histories will differ from one another: for example an inductivist internal history will differ from a conventionalist one, and both of these will differ from Lakatos's one. External history of science is the history of what actually happened, and it will take note of such non-rational causal factors as the social pressures on a scientist, accidents of communication, states of health and sickness, and so on. According to Lakatos internal history need not be true. He says (*BS*, p. 107) that a historian giving the (internal) history of the Bohr theory of the atom should include the postulation of electron spin in his account, since if Bohr had thought of this in 1913 it would have fitted naturally into his original account of his research programme. Kuhn (*BS*, p. 143) says 'that what Lakatos conceives as history is not history at all but philosophy fabricating examples'.

---

[1] Roger C. Buck and Robert S. Cohen (*eds*.) [1971]: *PSA*. 1970, *Boston Studies in the Philosophy of Science*, **8**. Dordrecht, Holland: D. Reidel Publishing Company, D.fl. 120., Pp. lxvi+615.

As a philosopher I feel like replying to Kuhn: 'Well, none the worse for that'. Kuhn goes on to object that since the 'internal history' derives from the philosopher's methodology, the philosopher can learn from it only what he puts in. He therefore claims that 'Lakatos' meta-methodological method is in danger of reducing to tautology' (*BS*, p. 141). I think that the method would be in such danger if Lakatos were trying to learn from history. However such an attempt to learn from history would be open to Popper's objections against historicism. So I interpret Lakatos as doing something which is not open to Kuhn's objections: I suggest that Lakatos is trying to learn from his examples in much the way in which philosophers normally learn from examples, and for which fictitious examples are as good as factual ones.

Why bother with history at all then? *In principle* it would seem that methodological discussion could be done by means of completely fabricated examples of theories, laws, *etc.* In practice this would not be possible because it would demand of the philosopher the sort of scientific imaginativeness which is possessed only by very creative scientists, and if a philosopher did possess this talent he would do more good by furthering actual science than he would by merely philosophising about science. Methodologists need examples from the history of science only because it is too hard to think up fictitious ones. It does not matter, therefore, whether the history is quite true. 'Internal history', as reconstructed by Lakatos, is surely the sort of history which provides examples of most methodological interest.

In *BS* Lakatos is interested in internal history from a rather different point of view. Here internal history is advocated as a historiographical weapon. To get the actual facts internal history has to be supplemented and corrected by the external history. Lakatos's historiographical idea is that we can judge a narrative in the history of science by the extent to which the narrative makes the actual history approximate to the internal history. I suppose that his idea rests on the assumption that other things being equal it is best to assume that scientists have acted in a rational manner. However I am no historian and shall therefore not consider further Lakatos's advice on historiography.

Let us return to methodology, and to the question of how much, for the purposes of methodological discussion, the internal history need bear any close relation to actual history. I have given a purely practical reason why the methodologist needs historical examples, rather than completely fabricated ones. Lakatos has another reason. He adopts a compromise between the two following courses: (*a*) the hubristic course of prescribing a methodology to scientists *a priori*; (*b*) the refusal to make independent value judgments altogether, so that methodologies are appraised solely by the extent to which they fit the actual practice of scientists. In *BS*, pp. 121 f., Lakatos compares *a priori* methodology with statute law and he compares the methodology which is abstracted from actual scientific practice with case law. Just as some compromise between pure statute law and pure case law is needed, so the methodologist must compromise between (*a*) and (*b*) above. I am reminded of the sort of compromise between utilitarianism and deontology which was made by such ethical writers as C. D. Broad and Sir David Ross. Lakatos holds that we should test methodological principles by measuring them against the practice of good scientists. However a methodology must not follow scientists slavishly: if it makes sense of most of the best work of the best scientists, it may also be used to correct the scientists too.

As Lakatos puts it 'The philosopher's statute law may occasionally be right when the scientists' judgment fails' (*BS*, p. 121) and 'The statute law approach should become much more important when a tradition degenerates or a new bad tradition is founded' (*BS*, p. 122). How do we make the appraisal that a scientist's judgment fails or that a scientific tradition degenerates? The appearance of circularity here is, I think, an appearance only. No methodology will make sense of everything that scientists do, not even if we restrict ourselves to scientists who are regarded as good scientists by all sorts of methodologists. However one methodology may imply that more of what scientists have done is reasonable than any other methodology does. This gives a reason to espouse this methodology, and then by the standards of this methodology we can appraise some of what scientists do as bad or degenerating.

Nevertheless, even with these qualifications, I find some difficulty over Lakatos's reliance on the intentions (or the practice) of the best scientists. What is the justification for such reliance? Is it that the intuitions of the best scientists have led to fruitful theories and research programmes in the past and therefore they will probably do so in future? Surely not, for to argue in this way would be to retreat both to inductivism and to historicism. In his [1972], as he has done earlier in his [1968], Lakatos does indeed plead to Popper for a harmless 'whiff of inductivism', but this is in the shape of a speculative (though criticisable) metaphysics which will justify the methodology which he extracts from his internal history of science. He leaves it unclear as to what this metaphysics will turn out to be like, and he also leaves it unclear as to the manner in which the metaphysics is to be criticised. If the metaphysics is to be accepted solely because of its ability to justify the methodology of the best scientists, then this leaves scientists as the ultimate umpires, and it is hard to see how such a justification of a methodology would have any better philosophical credentials than would a straight out inductivist one. (The only reasons for thinking otherwise would seem to be historicist ones.)

Probably I am still failing to understand Lakatos. He is at pains to say that his methodology is not a heuristics. His methodology is not meant as advice to scientists now and in the future but as an *ex post facto* appraisal of past science. (See *BS*, p. 174.) But here again I find a conflict with a Popperian doctrine which is congenial to me. What is the point of appraisal as such? Surely appraisal is valuable only if it is a guide to decision. In footnote 18 to Chapter 5 of his *Open Society and its Enemies* Popper remarks: 'But it is clear that moral *judgments* are absolutely irrelevant. Only a scandalmonger is interested in judging people or their actions...' Analogously, if Lakatos's methodological principles are not meant as heuristics, what is the point of them? What is the point of saying that a scientific research programme is a *good* one if this is not meant as advice to follow it or to do likewise? I shall return to this matter later when I discuss Feyerabend's criticism (in *CGK*) of Lakatos.

In his paper in *CGK* Lakatos expounds his methodology on the basis of a criticism of Popper's falsificationism, though his own positive and very original suggestions are very much within the general spirit of Popperian philosophy. According to Popper we use *modus tollens* in order to reject hypotheses, but such uses of *modus tollens* must not be thought of as *proving* that the hypotheses are false. According to Popper nothing is certain, and so logic leaves open the possibility of saving a hypothesis by modifying various assumptions. For example the testing of Newtonian mechanics depends on certain auxiliary assumptions about optics and about the existence or non-existence of invisible planets. Moreover even the observations used in the argument by *modus tollens* are fallible (all the more so, indeed, because observation statements are to a greater or lesser extent theory laden). Thus Popper ([1934]) holds that in order to apply *modus tollens* we need to adopt *conventions*: before we test a hypothesis we agree not to save it by making *ad hoc* changes in our auxiliary assumptions or by questioning the truth of our observation reports. We must not exaggerate this conventionalism, however, because Popper is quick to point out that we can on other occasions go on to test the auxiliary assumptions and the observation statements themselves. At any one time something must be accepted by convention, though nothing need always be so accepted. Moreover, in section 20 of his [1934], Popper allows that sometimes it may be proper to save a theory by means of a change in the auxiliary assumptions. This may be done when such a modification increases the empirical content of a theory, that is, when it augments the class of observational consequences. Lakatos holds, however, that Popper's falsificationism still leads to undesirable results. He points out that scientists have often hung on to a theory despite being unable to explain away anomalies (except in an *ad hoc* manner, that is in a manner which would not increase the empirical content of the theory).

Popper's falsificationism has to do with the demarcation between scientific theories and pseudo-scientific theories. Lakatos, however, holds that it is not individual theories which are appraised as scientific or otherwise, but sequences of theories, or 'scientific research programmes'. Popper had allowed that it is permissible to save a theory by modifying auxiliary assumptions, so that the total theory T becomes a slightly different theory T', provided that T' contains excess empirical content over T. Lakatos proposes instead that what we appraise as scientifically respectable or otherwise are certain *sequences* of theories ('scientific research programmes'). A scientific research programme is a good or 'progressive' one if later theories in the sequence contain excess *corroborated* empirical content over earlier ones. This is compatible with all the theories having false empirical consequences. These are *anomalies*, which have to be tolerated on pain of sterilising science completely. Eventually, it is hoped, of course, that these anomalies will be explained in a way which is not objectionably *ad hoc*, but on Popper's methodology pretty well all scientific research programmes would have been killed off before they got anywhere.

Lakatos calls a shift from a theory T to a later one T' in the sequence a 'problem shift'. The problem shift from T to T' is scientifically acceptable if the class of verified observational consequences of T is included in the class of verified observational consequences of T', while the class of verified observational consequences of T' is not included in the class of verified observational consequences of T. Typically, of course, the new theory T' will contain not only verified consequences which are anomalies relative to T but will contain new predictions which are subsequently verified during the period of the problem shift. Lakatos calls a problem shift of this sort a 'progressive' one, while a shift to a theory with fewer verified consequences is a 'degenerating' one. Lakatos's methodology is therefore highly consonant with the feeling which scientists have that a new theory is a good one if it predicts observations which had not been made at the time of acceptance of the theory which it replaces. The 'timeless' nature of justificationary methodologies made it impossible for them to account for the scientist's demand for the prediction of *not yet observed* phenomena. (See Lakatos *CGK*, p. 123.)

Because Lakatos holds that what is scientifically rational or otherwise are problem shifts within *sequences* of theories, his methodology implies a theoretical pluralism which goes beyond that of Popper who advocates a plurality of theories for merely heuristic reasons. (See Lakatos *CGK*, p. 121, footnote 4.) Feyerabend ([1962], pp. 65–6) has given other cogent reasons for theoretical pluralism which are also of a not merely heuristic sort. He has pointed out that the Brownian motion could not be seen as a counter-example to phenomenological thermodynamics save from the standpoint of statistical thermodynamics.

A sequence of theories must of course possess a certain continuity for it to constitute a single scientific research programme. Thus the proposal to explain the anomalous behaviour of the perihelion of Mercury by postulating an unobserved planet between Mercury and the Sun clearly belonged to the Newtonian research programme. So possibly did the development of electrostatics, since this involved little more than the addition of a new sort of central force which like gravitational force acted according to an inverse square law. The notion of a scientific research programme is not a very precise one. (Nor, I take it, was it meant to be a very precise one.) When Bohr modified the theory of central atomic forces by introducing quantisation, was this a departure from the classical programme (because of the quantisation) or was it merely a development of it (because of the central forces)? Similar questions about Kuhn's notion of a scientific revolution are raised by Stephen Toulmin in his contribution to *CGK*. There is clearly, as Kuhn recognises (see *BS*, p. 138) a close connection between Lakatos's concept of a scientific research programme and Kuhn's concept of a paradigm. However I take it that scientific research programmes get mentioned in Lakatos's 'internal history', whereas paradigms get mentioned in 'external history'.

In his witty paper 'Consolations for the Specialist' in *CGK* Paul Feyerabend claims that Lakatos's methodology gives no guidance to the scientist. (See *CGK*, p. 215.) This is because an initially degenerating scientific research programme may in the future become a uniformly progressive one, just as a curve may descend and then for ever ascend. Similarly an initially progressive programme may later become a degenerating one. A similar criticism is made by Richard J. Hall (*BS*, p. 152). In a note appended to p. 104 of *BS*, Lakatos replies to Feyerabend that methodology is meant to appraise scientific research programmes but is not meant to give heuristic advice as to whether or not to stick to such programmes. Lakatos remarks that it is perfectly rational to play a risky game. He says that what is not rational is to deceive oneself about the fact that a research programme has been degenerating. Hall's reply to this (*BS*, p. 152) is that honesty is not a sufficient condition for rationality. My own inclination is to return to a point which I made earlier. This is to ask what the point of appraisal of a methodology is if such an appraisal does not imply a heuristics. Can one ever divorce appraisal from recommendation? Suppose that I say that someone is a good lecturer. Presumably I do this in order to recommend people to go to his lectures or perhaps to appoint him to their university. Appraisal for its own sake, divorced from prescription or recommendation, is surely pointless. (In the case of moral appraisal it is even pharisaical. See the passage from Popper's *Open*

Society to which I referred earlier.) In any case Lakatos's remark that to follow a degenerating scientific research programme is to play a risky game is hard to reconcile with his divorce of appraisal from heuristics. How does Lakatos know that degenerating research programmes are risky? I take it that 'risky' here means 'not *likely* to become a progressive research programme'. But then to say that a policy is risky is to make a heuristic proposal. Nor would it do to replace 'risky' by 'possibly risky'. *Any* policy is *possibly* risky. A similar worry about Lakatos's intentions comes from a passage in his [1972], where he says that 'the value of excess corroboration is that it indicates that the scientists *might* be approaching truth.' But where does the *value* of excess corroboration arise here? 'Might' does not imply 'probably is'. A temporarily degenerating research programme *might* be approaching truth by a roundabout route.

If Lakatos's divorce of appraisal from heuristics can not be allowed, then even he can not quite evade the charge of historicism (despite his own disavowals of it). If a favourable appraisal of a research programme implies that it is likely to be progressive in the future, and if we say that this appraisal is tested by reference to the intuitions (or practice) of the best scientists in the past, then we *are* saying that the methods which worked in the past are likely to work in the future. (Lakatos often speaks of the *intuitions* of scientists where I would talk of their *practice*. The difference is unimportant, because presumably the intuitions are not meant to be a Platonic confrontation between the soul and reality, but some sort of inarticulate knowledge extracted from articulate practice.) But would not such an extraction of recommendations for the future from what has happened in the past be historicism? It is clear that to avoid historicism Lakatos *has* to divorce appraisal from recommendation, a course about which I have also found difficulty.

As I have already mentioned Lakatos keeps pleading to Popper for 'a whiff of inductivism'. This would be to propose a piece of speculative (but criticisable) metaphysics whose postulates would justify the favoured methodology. But it is not clear to me in just what way Lakatos envisages that this metaphysics would be criticisable. Simplicity? Occam's razor? But are not such principles of criticism as much in need of justification as are the inductive principle of inductivist philosophers or the methodological principles of Lakatos? It is hard not to sympathise with Feigl (*BS*, pp. 147–50), who still hankers after a vindication of induction or of methodology. (For Feigl's notion of vindication, as opposed to validation, see his [1950]). Another approach might be a Quinean one. Take methodology to be a theory about how scientists succeed, and use our existing scientific knowledge to support such a theory. (Cf. Quine [1969], p. 90 and pp. 126 f.) This approach probably works well for animal and low level induction (though it would not satisfy the philosophical sceptic, and is not meant to satisfy the philosophical sceptic), but it is hard to see how it could be applied to a philosophical pluralism such as that of Lakatos. There is therefore much unfinished (perhaps unfinishable) business on hand.

Let me briefly raise another difficulty which seems to arise for Lakatos and which is connected with his theoretical pluralism. The difficulty, which is raised by Kuhn on p. 266 of *CGK*, comes from the theory laden character of observation sentences, and Lakatos is presumably quite aware of it, though I do not know what his answer to it is. Suppose that T and T' are successive theories within a research programme. An observation sentence of T will be laden with the concepts of *T* and an observation sentence of *T′* will be laden with the concepts of *T′*. Presumably some of the concepts of *T′* might differ from those of *T* in important respects. For example suppose that we modify a pre-relativistic theory *T* so that it becomes a relativistic theory *T′*. (Surely, for example, the introduction of relativistic methods into quantum theory could count as a development within a single research programme.) Then, for example, 'has a mass in grams of' changes from a 2-adic predicate to a 3-adic predicate (it involves reference to an inertial frame). The observation sentences of *T* become different from those of *T′*. (Even Quinean doubts about synonymy could surely not impugn the non-synonymy of n-adic and (n+1)-adic predicates.) Perhaps this difficulty for Lakatos could be overcome by counting an observation sentence *O* of *T* as the same sentence as an observation sentence *O′* of *T′*, provided that *O* and *O′* were stimulus synonymous, or were uttered in the same objective situation by scientists suitably imbued with the theories *T* and *T′* respectively.

Though Lakatos writes excellent and readable prose, I found his contributions to *CGK* and *BS* slow to read and hard to understand. There is a tendency to relegate vital points to footnotes, and the proportion of footnotes to text is rather high. There is also a tendency to expect the reader to make very many back and forward references and comparisons, so that one is always flipping pages. There are also many (apparently vital) references to other books and articles. One wishes for a more self-contained presentation, less full of cross-references, footnotes, allusions to other works and so on. We should not wish to follow Gilbert Ryle's practice of banishing such things altogether, but there is a happy medium. I think that these stylistic idiosyncracies (which are characteristic also of many of Popper's writings, though not of his contribution to *CGK*) may slow the reader down, so that he forgets what has gone before and loses his bearings, and thus becomes liable to the misunderstandings of which there are so many accusations in the Popperian polemical literature. Where others have apparently fallen into grievous error I have little hope that I myself have avoided it. What Lakatos writes is so interesting that no one would *wish* to misunderstand him.

In this review I have been concerned mainly with Lakatos's papers in *CGK* and *BS* and with the criticisms thereof. A contributor to *BS* who I have not so far mentioned is Noretta Koertge who attempts to avoid Feyerabend's objections in *CGK* by strengthening Lakatos's methodology. Lakatos in his reply argues against such a strengthening. I might here mention that *BS* is a very large volume, and only a relatively small part is devoted to the problems of Lakatos's methodology. There are very many quite unrelated parts of it which contain papers of high value, for example the symposium on 'Theoretical entities in Statistical Explanation' by Greeno, Salmon and Jeffrey, and a paper on unknown probabilities by Hintikka. I was particularly interested in a paper by Clark Glymour on 'Theoretical Realism and Theoretical Equivalence'. But let us return to *CGK*. The majority of *CGK* is concerned with a discussion arising from Kuhn's opening paper 'Logic of Discovery or Psychology of Research'? There is a hard hitting reply to Kuhn by Popper himself. Popper is characteristically anti-historicist in his remarks against L. Pearce Williams. Pearce Williams had ended his own contribution by remarking that 'we need a lot more examples', and by quoting Bolingbroke 'history is philosophy teaching by example'. Popper says that 'the idea of turning for enlightenment concerning the aims of science, and its possible progress, to sociology or psychology (or, as Pearce Williams recommends, to the history of science) is surprising and disappointing'. The contributions by John Watkins, Stephen Toulmin, and Margaret Masterman, are all concerned in different ways with the Kuhnian notion of 'normal science'. There is a final reply by Kuhn himself.

J. J. C. SMART
La Trobe University

REFERENCES

FEIGL, H. [1950]: 'De Principiis Non Disputandum . . .?' in Max Black (*ed.*): *Philosophical Analysis*, pp. 119–156.

FEYERABEND, P. K. [1962]: 'Explanation, Reduction and Empiricism', in H. Feigl and G. Maxwell (*eds.*): *Minnesota Studies in the Philosophy of Science*, 3, pp. 28–97.

KUHN, T. [1962]: *The Structure of Scientific Revolutions.*

LAKATOS, I. [1968]: 'Changes in the Problem of Inductive Logic', in Lakatos (*ed.*): *The Problem of Inductive Logic*, pp. 315–417.

LAKATOS, I. [1972]: 'Popper on Demarcation and Induction', forthcoming in P. A. Schilpp (*ed.*): *The Philosophy of Sir Karl Popper.* Already published in German: 'Popper zum Abgrenzungs- und Induktionsproblem', in H. Lenk (*ed.*): *Neue Aspekte der Wissenschaftstheorie*, 1971, pp. 75–110.

POPPER, K. R. [1945]: *The Open Society and its Enemies.*

POPPER, K. R. [1935]: *Die Logik der Forschung* (Expanded English edition: *Logic of Scientific Discovery*, 1959).

POPPER, K. R. [1957]: *The Poverty of Historicism.* (Revised second edition, 1961).

POPPER, K. R. [1963]: *Conjectures and Refutations.*

QUINE, W. V. [1969]: *Ontological Relativity and Other Essays.*

FOR AND AGAINST METHOD[1]

Introduction.
1 *The Papers from the Conference.*
2 *Feyerabend's 'Against Method'.*
3 *Methodology of Psychology, Indeterminism and Miscellaneous Papers.*

Introduction.

The papers in the first half of this volume were all written in connection with a conference on Correspondence Rules which was held at the Minnesota Center for the Philosophy of Science in May 1966. Not all of the papers are primarily concerned with the problems of the semantics and structure of scientific theories —other topics discussed include inductive logic and theories of scientific method. This section contains papers by Feigl, Feyerabend, Hanson, Hempel, Hesse, Maxwell and Rozeboom.

The second half of the book is also a *potpourri.* Meehl has two papers on methodological problems in the social sciences; Radner and Meehl each give criticisms of Popper's discussion of indeterminism in *Of Clouds and Clocks*;

---

[1] Review of M. Radner and S. Winokur (*eds.*) [1970]: *Analyses of Theories and Methods of Physics and Psychology* (*Minnesota Studies in the Philosophy of Science*, 4). Minneapolis: University of Minnesota Press. $12.50. Pp. ix+441.

Gunderson writes on the mind-body problem; Demopoulos criticises Grünbaum's views on the conventionality of geometrical metrics.

There are some good papers on an assortment of topics in the volume but there seems no good reason why they should be bound together. The editors have provided name and subject indices but have omitted the valuable short abstracts which are found at the front of the other volumes in the Minnesota Studies series. Since many authors do not summarise their positions either in the introductions or conclusions of their papers, the omission of the abstracts is a very great loss indeed. One wonders if the public would not have been better served by having the long articles (*e.g.* those of Meehl and Feyerabend) published as separate monographs and the rest submitted to the journals. Some of the material would have profited from the scrutiny of referees.

1 *The Papers from the Conference.*

One of the authors in this volume claims that logical empiricism is dead. If so, it hardly died at the 1966 conference in Minneapolis. Nevertheless, there is a vague sense of anomie which haunts the first of this volume, and certainly the papers of Hempel and Feyerabend explicitly challenge some of the most basic presuppositions of the logical empiricist research programme.

Herbert Feigl's paper, 'The "Orthodox" View of Theories: Remarks in Defense as well as Critique', provides a sympathetic account of the 'layer-cake model' approach. (Feigl happily adopts the description which Feyerabend first used to make fun of the whole enterprise!) It is not always very clear what either the critics or defenders of the so-called 'layer-cake model' have in mind. However, I think that advocates of the layer-cake approach would generally accept something like the following suppositions (although they might express them somewhat differently):

(1) The Capillary Model of the Meaning of Theoretical Terms—cognitive content seeps up from the observational base.

(2) The Bootstrap Model of Confirmation—evidential support comes only from singular observation statements.

(3) The Sedimentation Model of the Development of Scientific Knowledge—although the methods of science may not be inductive, the knowledge we gain through conjecturing and testing grows upward from the soil of experience.

It will be convenient to group the papers in the first half of the volume according to these three areas. First, the papers primarily concerned with theories of meaning.

George Moore says in *Jumpers*[1] that at one moment in history the onus of proof passed from the atheist to those who wished to defend theism, '... quite suddenly, secretly, the noes had it'. Something similar has happened with the capillary model of the meaning of theoretical terms. With hindsight one is surprised that this approach should have seemed plausible for so long. Why should one expect to be able to give a linguistic specification of the meaning of the words used to designate the *new* entities or processes postulated by developing science? Might not we hope to discover (or dream of) radically new sorts of things which could not be characterised with the available vocabulary? Why should a Marco Polo or an Einstein be restricted by linguistic provincialism?

---

[1] Stoppard [1972].

Hempel's article, 'On the "Standard Conception" of Scientific Theories', contains a very clear, detailed criticism of the capillary model which is recommended to the reader. Here I will only discuss the alternative model which he provides. In a slightly earlier, related paper, Hempel expressed his disaffection as follows: '. . . it is misleading to view the internal principles of a theory as an uninterpreted calculus and the theoretical terms as variables, as markers of empty shells into which the juice of empirical content is pumped, through the pipelines called correspondence rules.[1]

Hempel argues that instead of analysing a theory (T) in terms of a formal calculus (C) and correspondence rules (R), we might better view it as consisting of internal principles (I), which specify the theoretical entities and the laws that they obey, and bridge principles (B), which link the internal principles to the previously examined phenomena which are to be explained. The internal principles will be written in a new theoretical vacabulary (V_T) while the bridge principles will contain antecedently understood vocabulary (V_A) as well.

At first, the move from $T=(C, R)$ to $T=(I, B)$ may not look very significant. However, there are at least two important distinctions between Hempel's view and the so-called 'standard' one. First, the antecedent vocabulary is not restricted to observation terms; it will typically contain terms from earlier theories (p. 144). Secondly, the 'new' theoretical terms are not totally dependent on the bridge principles for their interpretation. 'For the theoretical scenario is normally described in part by means of terms that have a use, and are understood, prior to, and independently of, the introduction of the theory' (p. 153). We see that although the V_T vocabulary is 'new' in the sense that it has not previously been used in well tested, explanatory scientific theories, it is not incomprehensible. (As an example, Hempel discusses the interpretation of the terms 'atom' prior to the time when the kinetic theory was introduced to explain the behaviour of gases.)

Hempel's model gives a somewhat more natural account of the structure of scientific theories and I think that most of his criticisms of the calculus-plus-correspondence-rules approach are well-taken. But it should be emphasised how many of the original scruples of the logical empiricists he ignores. On Hempel's account we now have no demarcation between 'cognitively significant' expressions and nonsense, because it is claimed that we can understand vocabulary before it is used in empirically confirmable statements. Neither do we have an account of how non-observational terms gain their meaning—we do not even know whether the process of linking them with observational terms is involved. Hempel's account 'does not presuppose the analytic-synthetic distinction' (p. 161), and he now seems opposed even to talking about 'the notoriously vague and elusive notion of meaning' (p. 154).

But if we grant Hempel, as I am happy to do, that these traditional problems are 'misconceived' (p. 162) (thus it is no mark against his theory that it ignores them), then we must ask which problems his theory does solve! Hempel claims to be concerned with the logical structure and the epistemic status of scientific theories. But his theory seems to shed little light on either issue. He admits that the distinction between internal principles and bridge principles is vague and that the difference is *not* one of epistemic status (p. 161). The same seems to be true of the distinction between the V_T and V_A vocabulary items. Why, then, is either distinction useful or interesting?

---

[1] Cf. Hempel [1969], p. 31. I recommend this earlier paper to the reader. In it Hempel is blunter and his position is less qualified.

Grover Maxwell, in ‘Structural Realism and the Meaning of Theoretical Terms’, tackles the problem of theoretical terms by combining Russell's theory of knowledge-by-acquaintance *versus* knowledge-by-description and Ramsay's method of dealing with theoretical terms. His basic position, as I understand it, is this: we can have knowledge about unobserved entities (such as an atom which disintegrated or a man who stole a car in the dead of night) and even refer to them. However, in both cases this is knowledge-by-description, not knowledge-by-acquaintance. The Ramsay sentence of a scientific theory succeeds in referring to a theoretical entity through logically connected observational terms which describe its structural characteristics.

Maxwell's extension of Russell's approach is an interesting one although it will not be acceptable to anyone who is unwilling to substitute the Ramsay sentence for the original theory. It also leaves open the question of what sorts of objects one can know by acquaintance. Maxwell's own view, which he does not discuss in detail in his paper, is that the only observables are the ‘inner events of observers’ (p. 181). Both chairs and quarks are inferred entities, known only by description.

Rozeboom, in ‘The Crisis in Philosophical Linguistics’, also adopts the structural realist position and goes on to argue that there are not enough connections between observational expressions and theoretical terms to permit us to individuate theoretical entities; ‘. . . theoretical expressions do not, in general, have unique referents’ (p. 217). He then concludes that it is essential to develop a semantic theory for terms having multiple designations (p. 218).

I found many of the points in the paper very confusing. For example, at the end Rozeboom calls for a ‘logic of genuine multiple designation’ in which the following case may arise: ‘. . . a sentence $P(a)$ is, presumably, true if there exists a property designated by P which is possessed by some designatum of (a), so that $P(a)$ and $Q(a)$ can each be true separately even while their conjunction is false . . . (p. 218), ‘. . . the logic required is altogether different from the logic of common names [he is referring to systems with Leśniewski's epsilon operator] (p. 218).’

Hanson's paper has the very misleading title, ‘A Picture Theory of Meaning’. This is not to be confused with his paper of the same name in Colodny [1970] which *does* deal with picture theories of meaning. In the paper under review Hanson points out that in the development of science sometimes the formalism comes before the application of it to practical experimental situations (*e.g.* the case of Euler's and Bernoulli's theory of ideal fluids). Other times one has a lot of correlated data and the problem is to design a simple mathematical theory from which they can be derived (*e.g.* the case of aerodynamics). No very noteworthy philosophical point transpires but Hanson always had interesting things to say about airplanes. Let us now turn to the contributions of the conference to the problem of confirmation.

The bootstrap approach to confirmation theory is certainly very plausible if one is an empiricist. What sort of evidence *could* one bring forth in support of a theory other than that provided by sense experience? Yet all attempts thus far to provide foundations for theoretical science through a logic of confirmation have ultimately had recourse to sky hooks—assumptions about prior probabilities, families of basic predicates, symmetry, or limited variety. In 'An Inductive Logic of Theories' Mary Hesse claims that neither the question of what a theory means nor how it comes to be confirmed can be understood without analysing the important role which models and analogies play. Most of Hesse's present paper (which is a continuation of previous studies on the use of analogies in science and in the logic of science[1]) is about the following problem: Consider a theory which successfully explains a vast number of experiments. (Call their conjunction e_1.) Suppose that the theory also makes a new prediction (e_2). In at least some cases, such as at the time just prior to the explosion of the first atom bomb, we are inclined to put a good deal of confidence in the untested prediction. Yet 'there is a perfectly general argument which shows that no probabilistic confirmation theory of any type yet developed will allow us to infer with higher than prior confirmation from e_1 to e_2 *merely in virtue of the fact that both are deductive consequences of some theory* . . . (p. 165). Hesse then gives a detailed discussion of how the degree of confirmation can be raised if the theory is confirmed not only by its entailments but also through an external model. 'Theories cannot be pulled up by their own bootstraps but only by the support from external models' (p. 175). *But what supports the external models?* Since the model which gives external support includes another rather reliable theory (or the same theory tested in another domain), we must ask how the theory in the model received a high degree of confirmation.[2] If the theory in the model gains its confirmation from another external model, we are led to a regress. (One also wonders if the meta-statement to the effect that the theory in question and the external model are sufficiently analogous to provide the basis for confirmation is not really an ordinary scientific theory in disguise. If so, how does the super-theory regarding the strength of the analogy get confirmed? On the other hand, if we decide to measure the strength of the analogy by merely counting the number of shared predictions then the measure is very sensitive to our choice of basic language. Once again we seem to need a sky hook—perhaps a doctrine about natural analogies.)

Part I also contains a comment by Professor Margolis, 'Notes on Feyerabend and Hanson', and forty pages of discussion from the conference. The editing of the discussion is unfortunate. At first glance one might assume that it is essentially a verbatim transcription with a little cutting. There are the redundancies and non sequiturs which one expects in discussions but which make for annoying reading. Moreover, some participants were very unhappy about the way their comments were reported. In general, both live discussions and the written version of them would profit, I should think, if the participants were allowed to provide their own rational reconstructions of their contributions. Historians will have to look elsewhere than in the professional literature for a literal transcript of the proceedings.

If we are to judge from the discussion reports, the real focus of interest at the conference was not any issue arising out of the papers but the wretched incommensurability thesis! The doctrine of the theory-ladenness of observation statements is an interesting one which would present a real challange to the logical empiricist approach if it could be made reasonably precise. Unfortunately most recent versions of this view have been so overstated that any first year student can easily attack them. By endorsing an extreme holism and claiming to find radical meaning change everywhere the incommensurabilists effectively nullify their legitimate criticism of the layer-cake model.

---

[1] See Hesse [1965] and [1968].

[2] That it does have a high degree of confirmation is indicated by a passage in which Hesse suggests that the theory of the model (t^*) be included along with e_1 as part of the evidence (p. 175).

The 1966 discussion, as reported here, is a quite typical case. Hanson opens the discussion with a remark about different conceptual schemes. Hempel replies: 'The problem as I see it is this: "What criterion can one use for deciding when the meaning of a term changes?" ' (p. 221). No answer is forthcoming, of course, and the circus begins. For example, when Hanson is asked if he really meant to say that the Wright brothers did not understand what the word 'flight' meant, he replies: 'There is a sense in which if you simply are sitting there, fat, dumb, and happy, and a bird takes off from under your nose, you can say of it that it is in flight now. Nonetheless . . .' (p. 243). I should add that there are some interesting comments by people who did not read papers at the conference, including Achinstein, Buck, Craig, Grünbaum, McMullin and Salmon.

2 *Feyerabend's 'Against Method'*.

As the title may suggest, Paul Feyerabend's paper,[1] 'Against Method: Outline of an Anarchistic Theory of Knowledge', is an iconoclastic romp which certainly adds spice to an otherwise staid volume. For one example of his eloquence and wit, be sure to read the long footnote attacking the sterility of the style adopted in technical papers, for which 'the awful Newton' is blamed. There he quotes two dreadful samples of scientistic language, one taken from *Human Sexual Response*, by Masters and Johnson, and the other from a Berkeley student pamphlet (pp. 95–9)! Feyerabend even manages to liven up the indices. Over half of the entries on the first page of the name index come from his article—few other contemporary philosophers would refer to Agrippa, Aristophanes, Bakunin, Isaac Barrow, and Bede. And in the subject index we find marvellous entries such as ' "Anything goes," principle of' and '. . . Dadaism'.

As usual I find it difficult to tell which of Feyerabend's comments to take at face value and which are attempts at that 'heavenly maliciousness' which Einstein found so delightful in Galileo.[2] (Probably Feyerabend would be delighted at my distress because it may indicate that he has successfully mixed propagandising with case making!) For example, many of his comments on science are interspersed with quotations from Marxist-Leninist-Maoist political thought. At first this looks like just an amusing diversion. Many of the fragmented quotations are rather badly torn out of context[3] and it is fun to compile one's own list from authors whom Feyerabend might not admire so much. One might suggest a quote from Mussolini as a motto for the section on using counter-induction to proliferate theories: 'I do not exclude [these ideas]. I never exclude anything. Anything is possible, even the most impossible and most senseless.'[4]

---

[1] Feyerabend was at the conference but this paper was submitted later.

[2] See Einstein's introduction in Baumgardt [1951].

[3] For example, see the rest of the paragraph from which the Hegel quotation given on the first page of the essay is taken.

[4] Cited in Mannheim [1936], p. 136.

However, as the paper progresses, we see that Feyerabend's references to social and political philosophers are not completely gratuitous. He would very much like his theory of science to grow out of a general theory about man and society.[1] So even though the paper is called 'Against Method' and most of the sections are directed against the methodological consequences of the Popperian philosophy of critical rationalism, Feyerabend's main concern is to challenge the received theories about the aim and value of science. Thus after a lucid summary of Popper's theory of scientific development[2] Feyerabend asks two questions, the first being of preemptive importance:

> '1. Is it *desirable* to live in accordance with the rules of a critical rationalism?
> 2. Is it *possible* to have both a science as we know it and these rules?' (p. 76).

Feyerabend's answer to both questions is 'No'. Critical rationalism fails both as a characterisation of effective scientific practice and as a philosophy of life. I will discuss his views on these two questions in reverse order.

In this paper Feyerabend not only rejects various prevailing views of scientific method but also takes a strong stand against the attempt to articulate *any* universal rules of rational scientific procedure:

> '[For any rule] there are always circumstances when it is advisable not only to ignore the rule, but to adopt its opposite. For example, there are circumstances when it is advisable to introduce, elaborate, and defend *ad hoc* hypotheses, or hypotheses which contradict well-established and generally accepted experimental results, or hypotheses whose content is smaller than the content of the existing and empirically adequate alternatives, or self-inconsistent hypotheses, and so on' (p. 22).

Feyerabend then goes on to strengthen and augment some of his earlier arguments against what he takes to be the received views about good scientific practice. Many of his criticisms are excellent—I will summarise a few of them in a moment. But first let me emphasise that Feyerabend presents no explicit argument for his very strong claim that there can be *no* rules which describe good scientific practice, although he frequently states this.[3] Perhaps he is arguing by induction since all the methodological maxims thus far proposed have been faulty! Even if one held a historicist theory whereby the laws of thought were believed to change, it should at least be possible to describe procedures which were valid for a given historical epoch.

Although Feyerabend recommends that we proceed counter-inductively and insists that the only methodological principle which always holds is 'Anything goes' (p. 26), he *does* advise the prospective scientist to study episodes in the history of science. However, this should be done 'with a novelist's love for character and detail... rather than with the crude and laughably inadequate instruments of the logician' (p. 19). This practice will 'develop the imagination of the student' and he may also pick up 'rules of thumb, useful hints, and heuristic suggestions...' (p. 18). Feyerabend even admits the possibility of developing a 'theory of error' which would describe some of the ways in which error has entered into scientific investigations in the past (p. 18). Unfortunately, he does not develop this idea further.

---

[1] I maintain this despite Feyerabend's explicit denial in the first sentence of his paper that his position assumes that anarchism is the most attractive *political* philosophy (p. 17). (My L.S.E. informants tell me that this disclaimer was added in proof. In any case, this isolated sentence is not consistent with the rest of his paper.)

[2] I have found this short section, 'Rationality Again' (pp. 72–81), a useful one for beginning students. They particularly enjoy one of the opening sentences: '... let us look at the standards of the Popperian school with whose ratiomania we are here mainly concerned' (p. 72).

[3] 'My discussion of Galileo ... has the aim of showing that such a 'correct method' does not *and cannot* exist' (p. 55).

Here are summaries of some of Feyerabend's criticisms of common methodological rules:

(1) It is unreasonable to demand that any new theory deductively explain its successful predecessors (or even an approximate, qualified version of them), because often scientific progress requires the invention of radically new concepts. Feyerabend has tried to make this point before, but the discussion of the incommensurability of the Galilean and Aristotelian concepts of notion in this paper (pp. 55–61) is by far the clearest and most persuasive example of radical change in meaning which I have seen.

(2) Feyerabend claims that a good reason for working with theories which are inconsistent with observation reports is the fact that the evidence may be contaminated with older, unacceptable views (p. 44). He then discusses 'natural interpretations' of phenomena and how they may be changed (pp. 48–55). I find this discussion of 'theory-ladenness' much better than most.

(3) Not only does every theory spend its entire life in a Lakatosian 'sea of anomalies' (*i.e.*, there are numerical disagreements between theoretical prediction and experiment), often there are also 'qualitative failures', as Feyerabend calls them. Yet progress has often been made by ignoring these gross inconsistencies.

For instance, 'classical electrodynamics contains the absurd consequence that the motion of a free particle is self-accelerated' (p. 39). As another example Feyerabend points out that, according to Newton's corpuscular theory of light, the somewhat rough surface of a mirror should scatter the corpuscles in all directions instead of providing clear reflections (pp. 39–40). There have been other discussions of examples similar to the ones Feyerabend gives[1] and I think that the problem of providing a general theory of the role of such cases in the growth of science is a pressing one. There have been some cases, such as the Newton example, where the qualitative failures of a theory have been largely ignored. But there are other situations (*e.g.*, the discovery of the rare-earths, which Soddy describes as 'breaking' the Periodic Table) in which they have played an important role in the historical development of a scientific theory.[2] Can one give a criterion which distinguishes between the two cases? Was Newton justified in largely ignoring the problem about the reflections of mirrors? Or, as Feyerabend would suggest, is it impossible to say?

(4) Feyerabend also criticises Popper's strictures against *ad hoc* hypotheses and claims to have found an ally in Lakatos concerning this matter (p. 64). Although Lakatos will probably set the record straight himself in the forthcoming new edition of 'Against Method' (to be published by the New Left

---

[1] Post [1971] describes them as 'flaws' and thinks that they can serve as heuristic guides to better theories.

[2] I discuss cases where 'striking counter-instances' have played an important role in the rejection of a research programme in my [1971] and [1972].

Publishing House with a commentary by Lakatos), I should perhaps point out here that Lakatos's theory does not espouse the use of *ad hoc* hypotheses. Quite the contrary. Lakatos distinguishes three senses of *ad hoc*.[1] Theories may be *ad hoc* because (*i*) they do not add content to existing background knowledge, (*ii*) they fail to make any novel predictions which are correct, or (*iii*) they do not have the internal unity provided by a 'positive heuristic'. Lakatos says all three senses are 'unfailingly perjorative'[2] and claims that his methodology is better than Popper's because it rules out *ad hoc* stratagems of the third type.[3] (It is this move which permits Lakatos to criticise much of current research in social science.) I did not find Feyerabend's discussion of Galileo's purported use of *ad hoc* hypotheses (pp. 64–9) as persuasive as his other historical examples.

Let us now turn to the question Feyerabend is most interested in: Even if one could best advance science by following rules of rational scientific practice (which he denies), *ought* one to do so?

In earlier papers, Feyerabend has often affirmed his belief that 'the happiness and the full development of an individual human being is . . . the highest possible value'[4] and has criticized various philosophies of science for beng anti-humanitarian.[5] And in this paper he praises Mill for putting the welfare and development of the individual as the primary aim and for making sure that his characterisation of scientific method would facilitate this aim.

Although he has long emphasised questions about the direct human consequences of adopting a certain scientific method as well as considerations of their effectiveness in advancing science, Feyerabend has generally been optimistic about the possibility of pursuing happiness and scientific truth at the same time. For example, theoretical pluralism was recommended not only because it was essential for the growth of science, but also because it might serve as an antidote to psychological dogmatism and institutional rigidity. Most importantly, Feyerabend felt that it was kind of methodology appropriate for a free individual to use.

However, in the paper under discussion Feyerabend seems to find less compatibility between the search for scientific truth and the pursuit of happiness: 'It is good to be constantly reminded of the fact that science as we know it today is not inescapable, and that we can construct a world in which it plays no role whatever. (Such a world, I venture to suggest, would be more pleasant to behold than the world we live in today, both materially and intellectually)' (p. 90). It now appears that Feyerabend is suggesting that those who love both science and the good life may be faced with some hard choices.

Many people who share the general concern about the quality of life in our technological society suspect that the present emphasis on, or misuse of, science may be a major contributing factor, but few have given as radical an analysis of the cause of our present difficulties as the one presented in this article. As I understand him, Feyerabend thinks that the scientific ideolgy is harmful because it is preoccupied with rationality in general, and argumentation and the discovery of methodological rules in particular. Feyerabend is not just attacking perverted forms of rationality, such as governments' attempts to 'rationalise' genocidal policies. He suggests that not only is argument generally ineffective (a change of consciousness or the adoption of a rival faith is often needed), but also that the process of coming to be persuaded by rational arguments is itself harmful: '... rational arguments work with rational people only. ... Rational people are specially prepared, they have been conditioned in certain ways, their freedom of action and of thought has been considerably restricted' (p. 101). Feyerabend then goes on to endorse 'the unreasoning impulse of human kindliness' (p. 102).

---

[1] See Lakatos [1968], p. 389 and Lakatos [1970], p. 175.

[2] Lakatos [1970], p. 195.

[3] *Ibid.*, p. 182.

[4] See the section entitled 'A plea for hedonism' in Feyerabend [1970].

[5] In Feyerabend [1962], he claimed that the requirements of deducibility and meaning invariance would make scientists like the followers of a tyrant. He also considers Kuhn's normal science to be incompatible with a humanitarian outlook. See Feyerabend ([1970], p. 210).

It would be too depressing to draw up a list of the evils done in the name of well-meaning, but ignorant, human kindliness and compare it with those resulting from carefully argued, but inhumane, policies. Surely we do not have to choose between such unacceptable alternatives.

What would a Feyerabendian post-rationalist society be like? Feyerabend does not present us with a general social or political theory—the most we get are remarks suggesting that 'a combination of Mill's general ideas and of a practical anarchism such as that of Cohn-Bendit' would be the best starting point (p. 108). There are strong arguments against system building in the area of political philosophy, but we might at least expect him to deal with some of the obvious problems that arise from the juxtaposition of liberalism and his brand of anarchism. For example, Feyerabend criticises Marcuse for wishing to use power to suppress retrogressive elements rather than using the same power to provide for better education 'in the art of critical thinking' and the opportunity for 'a more balanced discussion of views' (p. 109). This all sounds highly acceptable to the traditional liberal. But elsewhere (in the sections where Feyerabend is arguing for the use of propaganda), he stresses the importance of indoctrination in education and claims that when arguments are effective this must often be ascribed to 'their *physical repetition*, not their *semantic content*' (p. 22). Are the citizens of the society Feyerabend advocates to be *indoctrinated* into the art of critical thinking? (This conjures up images of science students being exposed to repeated loud speaker announcements of: 'Remember our motto: Anything goes!')

We gain some clue to the details of Feyerabend's own position if we examine his criticisms of Popper's social-political philosophy. Feyerabend's first criticism is that Popper obtained his ideas about social and political philosophy by generalizing his solutions to certain methodological and epistemological problems. The events surrounding Popper's *discovery* of his theory of the open society need not concern us here.[1] However the problem of the logical relationship between Popper's theory of science and his theory of the open society is an interesting one. It is certainly true that much of Popper's social and political philosophy is based on his epistemological doctrines concerning fallibilism and criticism. However, he also stresses that men should not try to reconcile their differing views on what is valuable in life by using the techniques which are appropriate in cases of scientific disagreement. (One of his objections to Utopianism is its requirement that all must agree to a detailed description of what would make them happy.) Also it is striking how many tenets of Popper's philosophy of the natural sciences are modified when he treats the social sciences. Instead of the exhortations to test our theories severely, we are cautioned to rely on piecemeal engineering. In the natural sciences any sort of theoretical entity may be invoked as long as the theory is testable; in the social sciences Popper is a committed methodological individualist. We are told that theories in the natural sciences often arise out of myth and metaphysical speculation; but social scientists should always start with practical problems. The presumptive question is whether the details of the views in Popper's *Open Society* are consistent with the rest of his philosophy, not whether they were derived from it!

---

[1] For Popper's own account see his preface to the second edition of his [1945].

Let us turn to Feyerabend's main objection, which is that Popper does not sufficiently stress the importance and value of the individual: 'I recommend Luxemburg because in elaborating her method she has always the individual before her eyes (one cannot say the same about Sir Karl Popper)' (p. 105). Elsewhere he says, 'For Popper the search for the truth seems to be much more important [than the material and spiritual welfare of the individual] and it seems occasionally to even outrank the interests of the individual' (p. 110). Feyerabend does not provide us with any textual evidence for this claim. I think that the very most one could argue is that Popper did not sufficiently ***emphasise*** the extent to which his theory of the social sciences depended on humanitarian considerations as well as epistemological ones. For example, I would suggest that the major reason for Popper's recommendation that we restrict our social experiments to simple piecemeal corrections of social evils instead of major revolutions is that he thinks that small, controlled modifications are likely to be less dangerous to individuals. Popper also presents an argument about the causal chains between modification and result being easier to trace if only a few variable are changed (and by small amounts), but I do not find this methodological argument very persuasive. Watkins [1972] points out that one of the reasons for using piecemeal engineering is that we may be able to reverse the policy if the experiment seems to be going badly. If we were only interested in extending our knowledge there would be no reason to stop the experiment before conclusive results were in.

There is little doubt that Popper's theory of the open society cannot be derived from methodological and epistemological considerations alone. It might be better to characterise his overall philosophy as one of critical, rational ***humanism.*** If this point be granted, would Feyerabend find Popper's general position acceptable? I think not. There are at least two major points of disagreement. First of all Feyerabend, unlike Popper, thinks that the very activity of being critical and rational may be dangerous in itself. He quotes Kierkegaard with approval: 'Is it not possible that my activity as an objective [or a critico-rational] observer of nature will weaken my strength as a human being?' (p. 76). An important question, but one which deserves a detailed answer which can be criticised in a rational way! I would argue that one cannot efficiently carry out Feyerabend's aim of providing for the welfare of the individual without reliable theories to guide the process. Even a hedonist needs to have true theories about what makes him happy and how to obtain it.

Secondly, Feyerabend and Popper probably have differing theories about the type of individualism which should be promoted. In *The Open Society and Its Enemies* Popper discusses various views of individualism and points out that (*a*) individualism does not imply egoism, but is compatible with altruism[1] and (*b*) an individualist can support institutions (*i.e.*, he need not be an anti-institutionalist personalist) as long as the institutions operate to help people and do not become ends in themselves.[2]

I would speculate that Feyerabend does not agree with Popper's stress on altruism. Feyerabend explicitly endorses 'ruthless opportunism' in the area of scientific methodology (p. 17). And there are some indications that opportunistic strategies are not to be confined just to the process of wrestling secrets from Nature.

It seems to me that, given Feyerabend's conception of science, his endorsement of Galileo's purportedly extensive use of propaganda implies an approval of egoistic individualism. My argument goes as follows: Since according to Feyerabend the primary aim of science is happiness, not truth, it would not be legitimate for Galileo to justify his use of propaganda by arguing that people would be enlightened if they understood his theory. Neither could he have argued convincingly that people would be happier if they thought the earth was moving. Therefore, on Feyerabend's assumptions, we must conclude that Galileo selfishly wished to force his views on to other people—he was on a 'power trip', as they say. Since Feyerabend praises Galileo's purported use of high-pressure techniques, I conclude that he must approve of this sort of theoretical imperialism. The slogan, 'Let a thousand flowers bloom', now seems to condone throwing dandelion seed on your neighbour's lawn.

Feyerabend's paper will certainly provoke a flurry of activity. Historians will criticise the details of his case studies, but in the process will gain a new perspective on past events. Philosophers will try to exorcise the very real threat of irrationality, but will also liberalise or modify their methodological principles so as to accommodate many of his examples. Perhaps co-optation is the sincerest form of flattery.

3 *Methodology of Psychology, Indeterminism, and Miscellaneous Papers.*

In the second half of the volume we find two papers by Professor Meehl on methodological problems in the social sciences. 'Some Methodological Reflections on the Difficulties of Psychoanalytic Research' is interesting because it honestly and openly faces up to a conflict which must trouble many clinical psychologists. Meehl obviously places considerable reliance on the kind of data which figure in existential phenomenological approaches ('. . . there are sessions where the material "fits together" so beautifully that one is sure almost any skeptic would be convinced . . .') (p. 407); yet he sees the need to set up ways of classifying and 'objectifying' the verbal behaviour of the patient. He concludes his remarks with a rather vague methodological prescription based on a quotation from Aristotle: ' "It is the mark of an educated man to look for precision in each class of things just so far as the nature of the subject admits." No more—but no less, either' (p. 416).

I find Meehl's other paper on methodology, 'Nuisance Variables and the *Ex Post Facto* Design', to be one of the more valuable ones in the volume. Here

---

[1] Popper [1945], 1, pp. 100–1.

[2] *Ibid.*, p. 268.

Meehl provides criticisms of a very popular research design which appear to be quite devasting.

The general problem is as follows: If we wish to measure the causal influence of *X* on *Y* and we suspect that *Z* influences both *X* and *Y*, we must correct for the effect of *Z*. Meehl holds, however, that 'the result of holding constant an identified nuisance variable *Z* will, in general, be to systematically ***unmatch*** pair members with respect to some fourth (unidentified) nuisance variable *W*' (p. 377).

For example, if we want to determine the influence of level of schooling on subsequent income, we would undoubtedly decide to take into account the effect of intelligence. However, Meehl points out that if one matches early school leavers and graduates pair-wise according to IQ (or some other measure of intelligence), one thereby systematically ***unmatches*** the pairs with respect to some other significant variable, such as degree of parental support. If one then matches for parental support too, then the matched pairs will then differ greatly in some other respect, such as the need for achievement. (If a stupid kid with no help at home manages to graduate, he must have had something else going for him! And that factor may very well help him get a good job.) Meehl concludes that '. . . rather than being a minor blemish on the *ex post facto* design, the likelihood of systematic unmatching represents a major methodological weakness which is likely to corrode the entire investigative enterprise' (pp. 378–9).

Meehl's discussion of this problem raises many questions which he does not comment on. How are we to proceed, given his criticism? If we are interested in questions such as the effect of the level of schooling on subsequent income would we be better off not matching for intelligence at all? Or should we try to exhaust all the nuisance variables? Or should we abandon this approach entirely?

Meehl also discusses other difficulties, including the problem of causal-arrow ambiguity. He emphasises the influence of tacit theoretical assumptions on the design of statistical experiments. Some may find the descriptive content of his examples concerning racial and ethnic differences controversial.

Professor Gunderson's very clear paper, 'Asymmetries and Mind-Body Perplexities', discusses the 'investigational asymmetries' which are supposedly encountered if we compare the process of coming to know about other minds with the process of coming to know about our own mind. He argues that, contrary to what is usually claimed, we are actually empirically ***underprivileged*** in our access to ourselves (p. 284). Gunderson analyses processes which he considers to be somewhat analogous, such as the manner in which we come to know what our own eyes look like. One interesting feature of his article is the use of Chomsky's phrase structure grammar in an argument about self-scanning scanners.

I will not try to summarize Demopoulos' article 'On the Relation of Topological to Metrical Structure'. The interested reader might wish to consult Grünbaum's paper in the December, 1970 issue of *Philosophy of Science* which refers to this article.

There are two papers in this volume on Popper's Compton Memorial Lecture, 'Of Clouds and Clocks: An Approach to the Problem of Rationality and the Freedom of Man'. Unfortunately neither author seems fully to have understood the interesting position which Popper developed there.

In 'Psychological Determinism and Human Rationality: A Psychologist's

Reactions to Professor Karl Popper's "Of Clouds and Clocks" ', Meehl attributes to Popper the thesis that 'psychological determinism is incompatible with human rationality' (p. 310). However, Popper devotes an entire section of that essay to the distinction between psychological determinism and physical determinism.[1] There he argues that the thesis of psychological determinism is so vague as to be perfectly compatible with physical indeterminism: '. . . a behaviourist may *at the very best* predict that, under given conditions, a rat will take 20 to 22 seconds to run a maze: he will have no idea how, by specifying more and more precise experimental conditions, he could make predictions which become more and more precise—and, *in principle, precise without limit*.'[2] Since no measurement can be infinitely precise, any thesis which says that every observable event has an observable cause is compatible with physical indeterminism.

Meehl argues against the thesis which interests him (although, as we have seen, it is not Popper's) by proposing a counter-example. He suggests that the following experiment could be carried out: An experimenter puts five pennies into a jar, lets the subject put two additional pennies into the jar, and then challenges or coaxes the subject to believe for 10 seconds that there are now *eight* pennies in the jar. (Perhaps the subject is to be given ten thousand dollars if he is successful.) We then 'observe' the subject's beliefs by using a psychogalvanometer and find that he cannot suspend his belief that there are seven pennies in the jar.

Meehl then argues that it is surely *rational* for the subject to believe that the laws of arithmetic apply in this situation, yet this belief is *determined* by the experimental situation: 'By placing 5 pennies in the box and having the subject place 2 pennies in the box, I *determine* his belief that it now contains 7 pennies, and I render it *impossible* for him to believe that there are 8. There is, I submit, little or no more "play" in this system than there is in the elicitation of a reflex from a spinal animal or the putting of a sugar lump into solution' (p. 316). Elsewhere he writes: '. . . the behavior is determined as completely as we determine the behavior of macro-objects in ordinary physics' (p. 315).

There are two obvious methods of criticising this suggested counterexample—one might argue either that the behaviour is not determined or that it is not rational. Let us pursue each approach.

The above remark about physics immediately shows that Meehl is not really talking about physical determinism, since, according to our present theory, even the behaviour of macro-objects is not exactly determined. However, I would also wish to argue that the behaviour of the human subject is much less determined than that of billiard balls. First, one might question the outcome of the experiment—might not some subjects learn to control their galvanic skin responses so that these could not be used to indicate what their beliefs are? In describing the experiment Meehl makes the assumption that the subject believes with 'as much certainty as we can generally reach about any empirical matter' that 'there is nothing phony about the construction of the jar' (p. 315).

However, there is nothing in the design of the experiment which *prevents* the subject from criticising or doubting these auxiliary hypotheses nor makes it unreasonable for him to do so. In fact, anyone who has heard about psychological experiments has good reason not to take anything which happens in such situations at face value! I conclude that, regardless of one's philosophical conception of determinism, one might agree that there are serious doubts that the outcome of believing there are seven pennies in the jar is completely determined by the experimental situation.

---

[1] Popper [1966], pp. 9–10.

[2] Popper [1966], p. 10.

This is not to deny that the subject's belief is *related* to his experience of seeing five, and then two more, pennies being put into the jar. In his *Of Clouds and Clocks*, Popper stresses that a universe operating by sheer chance is just as incompatible with the existence of free men behaving rationally as is a deterministic universe.[1] What is needed is some intermediate degree of determination—what Popper calls 'plastic control'. Thus rational agents are influenced by their environment (such as their information concerning the number of pennies in the jar); nevertheless, they are free to criticise their own beliefs, theories, methods, arithmetical systems, *etc.* As Popper says in discussing the extent to which musical compositions are determined, 'Mozart and Beethoven are, partly, controlled by their "taste", their system of musical evaluation. Yet this system is not cast-iron but rather plastic. It responds to new ideas, and it can be modified by new trials and errors. . . .'[2]

So the fact that the subject's belief about the number of pennies in the jar is controlled by the experimental situation in which he finds himself would not lead us to doubt the subject's rationality unless the control were 'cast-iron', as Popper describes it. But we have good reason to think that the control in this thought-experiment is 'plastic'; the subject might suspect that some sleight-of-hand was involved, or that he had incorrectly counted how many pennies the experimenter had put in, or even that he had forgotten his arithmetic facts! This last suggestion is not a very plausible one, but the fact that it is *possible* for the subject to believe it shows that we do not have a case of cast-iron control.

Let us now ask under what circumstances one might consider the subject's belief that there are seven pennies in the jar to be an irrational one. If the subject were categorically unwilling or unable to criticise the statement that there are seven pennies in the jar, then I think Popper would want to deny that his belief is a rational one. Popper argues elsewhere that the main difference between the ideas of a rational man and an irrational or mentally sick man is this: '. . . a mentally healthy person's beliefs are not incorrigible: a person in good mental health always shows a certain readiness to correct his beliefs [my translation].[3] Popper goes on to compare the engaged or committed person so beloved of existential philosophers to a madman: 'It could even happen that all of the fixed opinions of such a man were "adapted", in the sense that they happen to coincide with the best solution available at the time. But insofar as the man is committed, he is not rational . . . .'[4]

On Popper's theory of rationality, if circumstances somehow prevented us from keeping our belief about the number of pennies in the jar open to criticism, then that belief (even if true) would become irrational. An analogous claim could be made about rational actions: To blink one's eye when a gnat flies by is appropriate, adaptive behaviour, but it is not rational unless one is free to criticise the behaviour and perhaps decide to do otherwise.

I suspect that there are interesting criticisms of Popper's theory of rationality and its relation to determinism which could be made.[1] However, as I have indicated above, Meehl's argument misses the mark. His paper does raise some interesting issues; it also has an extensive bibliography which might be useful.

---

[1] Popper [1966], p. 13.

[2] Popper [1966], p. 31.

[3] Popper [1967], p. 149.

[4] *Ibid.* [My translation.]

Professor Radner's short article, entitled 'Popper and Laplace', touches on various topics in a rather superficial manner. He criticises Popper's paper on 'Interdeterminism in Quantum Physics and in Classical Physics'. It seems to me that Radner may not fully appreciate Popper's distinction between scientific determinism and metaphysical determinism, but the reader will want to pursue the arguments there in detail. Radner also suggests that some of Popper's proposals in *Of Clouds and Clocks* could be illuminated by using the statistical methods of Wiener's cybernetics: 'The point of comparing Popper and Wiener is to underscore the idea that Popper's problems can be attacked with mathematical tools as well as with simple metaphors' (p. 426). However, Radner evidently does not intend to claim that cybernetics actually provides us with a better understanding of rationality. After what appears to be an endorsement of Wiener, he adds: '... whether there is anything in the program for philosophers will have to be determined by a careful analysis of the arguments used by Wiener...' (p. 426).

It's a pity that the author did not at least begin such an analysis.

NORETTA KOERTGE
Department of the History and Philosophy of Science, Indiana University

---

[1] Some questions which arise immediately: What is the nature of plastic control? Are there causal connections between our experiences and our beliefs? If so, how free are we to criticise our beliefs? If it is answered that there is only a probabilistic connection between experiences and beliefs or beliefs and actions, one wonders if the propensity interpretation is to be applied here. If I have a propensity to act rationally, in what sense do I make undetermined choices?

REFERENCES

BAUMGARDT, C. [1951]: *Johannes Kepler: Life and Letters.*

COLODNY, R. G. [1970]: *The Nature and Function of Scientific Theories* (*University of Pittsburgh Series in the Philosophy of Science*, **4**).

FEYERABEND, P. K. [1962]: 'Explanation, Reduction and Empiricism', in H. Feigl and G. Maxwell (*eds.*): *Minnesota Studies in the Philosophy of Science*, **3**, pp. 28–97.

FEYERABEND, P. K. [1970]: 'Consolations for the Specialist', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*, pp. 197–230.

HEMPEL, C. G. [1969]: 'On the Structure of Scientific Theories' in *The Isenberg Memorial Lecture Series*, 1965–66.

HESSE, M. B. [1965]: 'The Explanatory Function of Metaphor', in Y. Bar-Hillel (*ed.*): *Logic, Methodology and Philosophy of Science*, pp. 249–59.

HESSE, M. B. [1968]: 'Consilience of Inductions', in I. Lakatos (*ed.*): *The Problem of Inductive Logic*, pp. 232–46.

KOERTGE, N. [1971]: 'Inter-theoretic Criticism and the Growth of Science', in R. C. Buck and R. S. Cohen (*eds.*): *PSA. 1970*, *Boston Studies in the Philosophy of Science*, **8**, pp. 160–173.

KOERTGE, N. [1972]: 'Theory Change in Science', forthcoming in G. Pearce (*ed.*): *Conceptual Change.*

LAKATOS, I. [1968]: 'Changes in the Problem of Inductive Logic' in I. Lakatos (*ed.*): *The Problem of Inductive Logic*, pp. 315–417.

Reviews

SPENCER BROWN, G. [1969]: *Laws of Form*. London: George Allen and Unwin. £3.00. Pp. xx+142.

Bertrand Russell said that 'in this book Mr Spencer Brown has succeeded in doing what is very rare indeed. He has revealed a new calculus of great power and simplicity. I congratulate him.'

This is generous, but I believe it will eventually be recognised that Mr Spencer Brown has done more than that. He has not invented an arbitrary new calculus, but that particular calculus which can let us see deeper into the nature of mathematics. I am not competent to judge whether his notation and development of the calculus are appropriate. What for me is crucial is that he starts mathematics at what is surely the most fertile point: at *the drawing of a distinction*. Is this not logically prior to, and more powerful than, any other mathematical procedure or concept, for example than a set? Asymmetrical relations are logically richer that the corresponding symmetrical relations.

It may be helpful to point out that what Spencer Brown has done is to develop a calculus based on what C. S. Peirce called an aliorelative relation, or one *implying diversity*, present when no term has this relation to itself. (See Russell. *Introduction to Mathematical Philosophy*, p. 32.) It is difficult to imagine a better starting point for mathematics, or indeed for a new epistemology, as one reviewer has suggested.

I find it natural to ask that the foundations of mathematics, *i.e.* its primary procedures, should be isomorphic with the primary processes of the human brain. Most, perhaps all, processes of sense perception can be reduced to the perception of a difference. Replace 'perception of a difference' by 'drawing of a distinction', and you have Spencer Brown's primary concept. Thus one may consider—though I think he does not explicitly assert this, it is implicit in his argument—that his calculus starts with what may prove to be the basic feature both of perception and of all reasoning processes. The validity of this interpretation is not at issue for the readers of this work, which may be studied and evaluated without it. But this interpretation provides the reason why, though no master of the algebra of logic, I wish to call attention to this book. Whatever may be its fate at the hands of mathematicians judging it by technical criteria, I regard it as a work of philosophical importance opening up a realm of enquiry too little explored as yet: the theory of differences, in this case of asymmetrical relations, which applied to physics becomes the theory of spatial asymmetry, of temporal succession, and of their relations.

Indeed I still consider, on re-examining this book after a two-year interval, that it is a work of genius. For the moment that is of course a personal judgment, though other readers have thought the same. Mr Spencer Brown tells us that this calculus has been directly applied to the representation and solution of at least one scientific problem: the design of certain electrical circuits. Here I can express no view as I have not studied the relevant material. But if the author's subsequent work demonstrates this by more explicit demonstrations, perhaps in other fields, then no doubt will remain that this slim volume, the fruit of ten years' work, initiates a major advance both in mathematics itself and in the mathematics of science. Physics, for example, needs greater clarity on asymmetrical relations.

In the meantime the aim of this review is to recommend to all interested, even to those who may find the calculus itself too novel to be easy, the careful reading of the introduction and notes. These contain valuable passages on Gödel's theorem, on the distinction which should be drawn between the creative process of proving a theorem from axioms and the mere demonstration of the consequences of postulates, on the role of contemplation in arriving at truth, and much else. One is aware of contact with a mind of high originality dedicated to a task as yet only partly achieved.

LANCELOT LAW WHYTE

GUNDERSON, Keith [1971]: *Mentality and Machines*. New York: Doubleday. $1.95. Pp. xviii + 174.

Gunderson's lucid and stimulating monograph discusses most of the main issues that have exercised philosophers in regard to computer simulation and artificial intelligence.

He begins with a useful historical survey of the earlier controversy on the man-machine problem, and in particular with a discussion of the views of Descartes and La Mettrie. He points out how this controversy, just like the present one, was provoked by new technological achievements (Vaucanson's puppets) and was interconnected with problems about the ascription of mental predicates to animals (can apes be taught a language?). He then considers some currently fashionable arguments for and against the view that machines can think.

On the 'pro' side Turing described an imitation-game, for three players—a man, a woman and an interrogator—where the object of the game is for the interrogator to determine, by question and answer, which of the other two is the man and which is the woman. Turing thought that instead of asking 'Can machines think?' we should ask some more precise question, such as 'How well can, or could, they play the imitation-game?' And his answer was to predict that in fifty years' time they will be able to play it at least as well as human beings. Gunderson's criticism of this is two-fold. First, he describes another game in which boxes of stones that fell on a person's toe might simulate either a man or a woman's stepping on it, so as to show that capacity to deceive should not be regarded as a criterion of mentality. Secondly, he points out that in a human being one can always infer from the possession of one such ability, *e.g.* the ability to simulate the opposite sex, to the probable possession of many related skills and abilities, *e.g.* the ability to act, to impersonate, to mimic animal-cries, and so on. But, objects Gunderson, this inference could not be made in the case of a computer.

The point scored by Gunderson here is rather an *ad hominem* one, since in principle there seems no reason why the abilities with which robots are programmed should not be interlinked in ways comparable to those found in human beings. More penetrating, however, is his distinction between programme-receptive features of human life and programme-resistant ones. Thus the simulation of cognitive skills and learnable abilities seems to present merely technical problems to the designer of robots. But the imitation of pains, emotions, motives *etc.* presents difficulties of a different order, in that the criteria for successful simulation are not so obvious. Gunderson believes (following Wilfrid Sellars) that Locke's distinction between simple and complex ideas is similar to his own distinction between programme-resistant and programme-receptive features, respectively. He rightly concludes, however, that those who do research in the artificial simulation of human beings should aim to think out a way of capturing the difference to which he draws attention, not that this difference is intrinsically resistant to artificial simulation.

Gunderson also pays some attention to the views of those who think that being programmed is inconsistent with having a mind. His argument against this position is that insofar as human behaviour can be construed in terms of the performance of well-defined tasks it is inherently rule-governed and harbours an inbuilt compatibility with being programmed. But, though that argument is effective against those who urge a straightforward inconsistency between being programmed and having a mind of one's own, it is not so effective against anyone who claims that the inconsistency lies between being programmed *by another* and having a mind of one's own. Obviously a normal man devises many of his own programmes or plans of action, whether consciously or not, and is also born with certain innate programmes for learning and problem-solving. It is quite reasonable to hold that such a man, *ipso facto*, has a mind of his own. But if this is what you hold then you must also grant that so far as the man's programme originates in another's mind he has not got a mind of his own. It is just when they will not behave as we want that we say of a child, or a cat, 'It has a mind of its own'. In our customary way of looking at things, as in the eyes of the law, the normal performances of a robot are the acts of its programmer and he must take responsibility for them.

Two qualifications need to be made, however. The first is that the situation becomes rather different if much of the detail of the robot's programme is programmed to be decided for it in a wide range of ways by the contingencies of its encounters with reality. If it thus builds up a quite idiosyncratic character, even its original programmer (though he might still be legally responsible) might be inclined to say that it had now got a mind of its own, since its developed rules of behaviour would not be predictable in detail from its original programme: they would have escaped the programmer's control.

Secondly, it is important to distinguish having a mind of one's own, or thinking for oneself, from behaving in ways describable by terms from the vocabulary of mind. Even the well-drilled man who never thinks for himself may at least do some kind of thinking: specifically, he may think as he has been programmed to think. So it is not sufficient to distinguish, as Gunderson does, between the criteria for calling x a person and the criteria for ascribing mental predicates to x. This is an indispensable distinction which enables us to rescue the commonsense view that cats have beliefs and intentions without having to

Noûs

Nihil philosophici a nobis alienum putamus

A Quarterly Journal of Philosophy
edited by Hector-Neri Castañeda
Philosophy Department, Indiana University
Bloomington, Indiana 47401

NOÛS is committed to the publication of first-rate essays and short discussions on philosophical problems regardless of the author's philosophical school or point of view. It includes critical studies of recently published books rather than reviews and critical summary surveys of studies as well as symposia dealing with specific philosophical problems.

A quarterly, NOÛS is published in March, May, September, and November each year. For *individuals* the subscription rates are: one year, $8.00; two years, $14.50; three years, $20.00. For *bona fide students*: $5.00 per year. For *institutions*: $12.00 per year. Add 15 cents per year to Canadian subscriptions and 35 cents per year to all other foreign subscriptions.

CONTENTS

Volume VI, Number 2, May 1972

CONTENTS

JOHN WALLACH—Belief and Satisfaction
THOMAS SCHWARTZ—Rationality and the Myth of the Maximum
RICHARD SHARVY—Euthyphro 9d-11b: Analysis and Definition in Plato and Others

Symposium on Properties and Relations
  MILTON FISK—Relatedness Without Relations
  REINHARDT GROSSMANN—Russell's Paradox and Complex Properties
  NINO B. COCCHIARELLA—Properties as Individuals in Formal Ontology
  ROBERT K. MEYER—Identity in Cocchiarella's T*
Publications Received

Order from
The Secretary, NOÛS
Department of Philosophy
702 Ballantine Hall
Indiana University
Bloomington, Indiana 47401

STUDIES IN HISTORY AND PHILOSOPHY OF SCIENCE August 1972

EDWARD MCKINNON Theoretical Entities and Metatheories

ROBERT J. BAUM The Instrumentalist and Formalist Elements of Berkeley's Philosophy of Mathematics

PHILIP QUINN Methodological Appraisal and Heuristic Advice: Problems in the Methodology of Scientific Research Programmes

EDWARD H. MADDEN and MENDEL SACHS Parmenidean Particulars and Vanishing Elements

Essay Review Marshall Clagett, **Nicole Oresme and the Medieval Geometry of Qualities and Motions** by Edward Grant

Reviews R. S. Woolhouse, **Locke's Philosophy of Science and Knowledge** reviewed by G. A. Rogers; John W. Yolton, **Locke and the Compass of Human Understanding. A Selective Commentary on the 'Essay'** reviewed by Margaret Osler; Marian Przelecki, **The Logic of Empirical Theories** reviewed by John M. Nicholas

THE
REVIEW OF METAPHYSICS
A PHILOSOPHICAL QUARTERLY
Edited by JUDE P. DOUGHERTY

Vol. XXV, No. 3 March 1972 Issue No. 99

ARTICLES
Gregory Vlastos *The Unity of the Virtues in the* Protagoras
Richard M. Gale *On What There Isn't*
CRITICAL STUDIES
Nathan Rotenstreich *On Lévi-Strauss' Concept of Structure*
Jonathan A. Weiss *The Justification of Punishment*
BOOKS RECEIVED
Joseph W. Sterling and Staff *Summaries and Comments*
PHILOSOPHICAL ABSTRACTS
ANNOUNCEMENTS

Individual Subscription $6.00 Student Subscription $4.00
Institutional Subscription $10.00

THE CATHOLIC UNIVERSITY OF AMERICA, WASHINGTON, D.C. 20017
Twenty-Year Cumulative Index, 1947-1967 is available for $3.00 a copy.

THEORIA

A Swedish Journal of Philosophy

Edited by GÖRAN HERMERÉN *and* KRISTER SEGERBERG

CONTENTS

VOLUME XXXVII 1971 Part 3

Helge Malmgren: Moore's concept "indirect apprehension"

Håkan Törnebohm: Durations and distances in time

Hans Kamp: Formal properties of "now"

Alasdair I. F. Urquhart: Completeness of weak implication

REVIEWS

THEORIA, which invites contributions in all fields of analytic philosophy, is published three times a year. Subscriptions and requests for back numbers should be sent to the publishers, Messers. C. W. K. Gleerup, Öresundsvägen 1, 222 38 Lund, Sweden. Annual subscription is Skr. 30:— (approx. 6 US $).

FOUNDATIONS OF LANGUAGE

International Journal of Language and Philosophy

Board of Editors: Morris Halle, (M.I.T.), Peter Hartmann, (Konstanz), K. Kunjunni Raja, (Madras), Benson Mates (Univ. of Calif.), J. F. Staal, (Univ. of Calif.) Pieter A. Verburg (Groningen), John W. M. Verhaar, (Djakarta)

Editorial Assistants: W. G. Klooster and H. J. Verkuyl.

CONTENTS OF VOLUME 8, No. 2

Gerald A. Sanders and James H.-Y. Tai/Immediate Dominance and Identity Deletion—P. W. Culicover/OM-Sentences—Pieter A. M. Seuren/Autonomous Versus Semantic Syntax—Catherine Ringen/On Arguments for Rule Ordering.

Discussion: A Note on Not-Hopping (Thomas Wasow).

Reviews: Paolo Valesio/*Strutture dell'allitterazione. Grammatica, retorica e folklore verbale* (G. C. Lepschy)—Yakov Makiel/*Essays on Linguistic Themes* (W. P. Lehmann)—Roderick A. Jacobs and Peter S. Rosenbaum/*English Transformational Grammar* (Richard Stanley)—L. A. Hill/*Prepositions and Adverbial Particles* (Bruce Fraser)—Rudolf P. Botha/*The Function of the Lexicon in Transformational Generative Grammar* (Paul G. Chapin)—S. K. Šaumjan/Problems of Theoretical Phonology (Karl Horst Schmidt)—Short Notices—Publications Received for Review.

Subscription price per volume of four issues Dfl: 115—(US $37.38), including postage. About six or seven issues are published yearly. Personal subscription price on request.

D. REIDEL PUBLISHING COMPANY / P.O. Box 17 - DORDRECHT - HOLLAND

INQUIRY An Interdisciplinary Journal of Philosophy and the Social Sciences

Edited by Arne Naess
and Alastair Hannay

Numbers 1-2 *Summer 1972* *Volume 15*

Editorial Statement	
Insanity and Responsibility	Herbert Fingarette
Mental Structure and Self-Consciousness	Brian O'Shaughnessy
Spinoza and Mental Health	Paul Wienpahl
Towards an Ecology of Consciousness	Frank Barron
Schizophrenia and Perception: a Critique of the Liberal Theory of Externality	James M. Glass
The Validity of Psychotherapy	B. A. Farrell
Review Discussions:	
Richard Wollheim: *Freud*	Frank Cioffi
Thomas Szasz: *The Manufacture of Madness*	Ronald B. de Sousa
David Robinson: *The Process of Becoming Ill*	Maurice Roche
Erwin W. Straus, Maurice Natanson, and Henri Ey: *Psychiatry and Philosophy*	A. Hannay

Published quarterly by Universitetsforlaget, P.O. Box 307
Blindern, Oslo 3, Norway

Annual Subscription: $10.00, £3·60 including postage

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY

Edited by MARVIN FARBER

JUNE 1972, VOLUME XXXII, No. 4

CONTENTS

I. ARTICLES
  Sensations: Perceived Sensible Qualities? by DAVID SELIGMAN
  Imperfect Knowledge, by JOSEPH AGASSI
  Transitive and Intransitive Modes of Sentience, by JOSEPH MARGOLIS
  Worldhood, by CLIVE INGRAM PEARSON
  Freedom in the Perspectivism of Ortega, by WILLIAM J. KILGORE
  What Austin Does with Words, by JOHN A. DINNEEN, S.J.
  The Noematics of Reason, by GARTH GILLAN
  Analytic A Posteriori Truth? by D. GOLDSTICK
II. DISCUSSIONS, REVIEWS
III. NOTES AND NEWS, RECENT PUBLICATIONS

Subscription prices $7.50 per year for individuals; $9.00 per year for institutions. Separate issues, $2.00 and $2.50 respectively. Postage and handling for back issues only. Papers submitted for publication will not be returned unless accompanied by a self-addressed stamped envelope, or equivalent postage.

A circular listing the main contents of the Journal since 1940 will be sent upon request.

Address all communications to

PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH
State University of New York at Buffalo, Buffalo, New York, 14226

DIRECTORIES OF PHILOSOPHY

The world's most comprehensive source of information on philosophy . . .

DIRECTORY OF AMERICAN PHILOSOPHERS, 1972-73
Sixth Edition, Edited by Archie J. Bahm

Information on the United States and Canada: 9300 Philosophers with Addresses, Fellowships and Assistantships, 2900 Colleges and Universities, 98 Philosophy Journals, 55 Philosophical Societies, 125 Publishers of Philosophy Books, Statistics of the Profession, Complete indices, 662 pages.

INTERNATIONAL DIRECTORY OF PHILOSOPHY, 1972-73
Second Edition, Edited by Ramona Cormier, Paul Kurtz,
Richard H. Lineback and Gilbert Varet.

Information on 91 countries, excluding the U.S. and Canada: 5000 Philosophers with Addresses, 700 Colleges and Universities, 240 Institutes of Philosophy, 175 Philosophical Societies, 470 Philosophy Journals, 690 Publishers of Philosophy Books, Complete indices, 451 pages.

Publication Date: February, 1972
Price of the two volume set: $35. (Individual Volumes: $20.)

Philosophy Documentation Center, Bowling Green State University
Bowling Green, Ohio 43402, U.S.A.

MIND

A Quarterly Review of Psychology and Philosophy
edited by Professor **DAVID HAMLYN**

JULY 1972

I Frege: Review of Dr. E. Husserl's *Philosophy of arithmetic:* translated by E. W. KLUGE
II The Binary Operation called "Material Implication" Soberly Understood: C. F. KIELKOPF
III An Intensional Interpretation of Truth-Values: A. R. ANDERSON
IV Describing: W. G. RUNCIMAN
V Intention and Foresight in Criminal Law: H. OBERDIEK
VI Discussions:
  Presupposition and Entailment R. H. KANE
  Mr. Roberts on Strawson N. NELKIN
  Professor Hintikka on Descartes' "Cogito" R. MITTON
  Searle on Descriptions S. BLACKBURN
  Professor A. Ross and Some Legal Puzzles J. RAZ
  On Alf Ross's Alleged Puzzle in Constitutional Law N. HOERSTER
  The Prisoner's Dilemma II J. SMYTH
  Non-Dogmatism and Ethical Paradoxes BRENDA COHEN
  The Ordinary Language Lattice J. SUZMAN
  The Limits of Statement Denial P. SWIGGART
  After-Images C. E. BURLINGAME
  Professor Owen, Aristotle, and The Third Man Argument J. C. DYBIKOWSKI
  A Note on Hare on Imagining Oneself in the Place of Others H. S. SILVERSTEIN
  True Sentences and True Propositions G. ENGLEBRETSEN
VII Critical Notices:
  On Certainty: by Ludwig Wittgenstein. A. PALMER
  Speech Acts: An Essay in the Philosophy of Language: by J. R. Searle. L. HOLBOROW
VIII New Books
IX Notes

Annual Subscription £1.25, post free; single copy 40p, plus postage

Published for MIND ASSOCIATION by
BASIL BLACKWELL 49 BROAD STREET OXFORD

Volume 23 Number 4 November 197

The British Journal for the Philosophy of Science

CAMBRIDGE UNIVERSITY PRESS

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editor: IMRE LAKATOS

Assistant Editors: COLIN HOWSON, JOHN WORRALL AND ELIE ZAHAR

Subscriptions

The *Journal* is published quarterly. The subscription price is £4·50 net in U.K. (U.S.A. $14.00) per volume post free; single issues £1·50 net in U.K. (U.S.A. $4.75) plus postage. The subscription is £3·00 (U.S.A. $9.50) for members of the British Society for the Philosophy of Science, who should order direct from the publisher. Orders may be sent to any bookseller or direct to Cambridge University Press, Bentley House, 200 Euston Road, London NW1 2DB. Orders from U.S.A. and Canada should be sent to Cambridge University Press American Branch, 32 East 57th Street, New York, N.Y. 10022, U.S.A. Second class postage paid at New York, N.Y.

Notes to Contributors

Articles, correspondence and books for review should be sent to the Editor, Professor Imre Lakatos, at the Department of Philosophy, London School of Economics, Houghton Street, Aldwych, London WC2A 2AE. Articles should not normally exceed 5,000 words, and should be accompanied by abstracts of about 100 words.

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Footnotes, which should be as few as possible, should be typed separately, double-spaced at the end of the text. They should be referred to by consecutive numbering throughout the text, in which they will be inserted in page proof. Bibliographical references should be listed alphabetically at the end. The style of references should follow current *Journal* practice.

Long quotations are printed in small type without quotes, and should be so marked in the typescript. Single quotes should be used to form the names of words and sentences, and double quotes for all other purposes. Diagrams, tables, and illustrations should be on separate sheets, with their desired position in the text clearly indicated. Symbolic formulae appearing as separate lines should be clearly distinguished in the text, and foreign letters (Greek, German, etc.) indicated in the margin. Material to be printed in italics, especially English letters used as symbols ('*A*', '*x*', etc.), should be underlined. On matters of style and layout, authors should consult CHAUNDY, T. W. *et al.* [1964]: *The Printing of Mathematics*, London: Oxford University Press.

Offprints of articles may be ordered when corrected galley proofs are returned. Fifty are supplied free ; the price of extra copies is given on the order form.

© Cambridge University Press, 1972

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the London or New York office of Cambridge University Press.

ISI Tear Service, 325 Chestnut Street, Philadelphia, Pennsylvania 19106, U.S.A. is authorised to supply single copies of separate articles for private use only.

Advertising

Particulars of advertising in the *Journal* may be obtained from Cambridge University Press, Bentley House, 200 Euston Road, London, NW1 2DB; or in U.S.A. and Canada, Cambridge University Press American Branch, 32 East 57th Street, New York, N.Y. 10022, U.S.A.

Back Volumes

Volumes 1-16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17-21 may be obtained from Cambridge University Press.

Brit. J. Phil. Sci. 23 (1972), 295–304 *Printed in Great Britain*

CALCUTTA UNIV CENT LIB

Newcomb's Paradox Revisited

by MAYA BAR-HILLEL and AVISHAI MARGALIT

This paper attempts to provide a solution to the Newcomb Problem, which was first presented in Nozick [1969]. The author suggested there a solution of his own, with which he admitted to being dissatisfied, and invited further comments that might 'enable [Nozick] to stop returning periodically to [the paradox]' (*op. cit.* p. 143). We found the paradox every bit as intriguing as Nozick did, and hope that our solution can restore his peace of mind.

Suppose you are playing a game with a Being whom you believe to possess extraordinary predictive powers. The game proceeds as follows: Before you are two boxes. In one you can plainly see $1,000. The other is covered, so you cannot see what it contains. But you know that the Being has put into it either a million dollars or nothing, depending on what he had predicted that you will do come your turn to play. You have a choice between two actions: taking what is in both boxes, or taking what is in the covered box only. You know, however, that the Being played his move as follows: if he predicted that you will take both boxes, he has left the covered box empty; if he predicted that you will take the covered box only, he has put a million dollars in it. You are not allowed to use a chance device to determine your choice. You have enormous confidence in the Being's ability to predict your actions, and you know that he has correctly predicted all choices of all players who have played this game with him to date (in fact, he has predicted correctly all choices that *you* have made in this game in some previous 'warming up' trials, played for points, say, rather than money). The Being has just now (or an hour ago, or a year ago . . .) made his prediction and played his move. It is now your turn. What will you do?

Both actions can be argued for very persuasively and intuitively.

An argument for taking only the covered box might run as follows: You firmly believe that whatever you ultimately decide to do, the Being has probably foreseen. In other words, you are, for some reason, almost sure that if you will take both boxes you will end up with $1,000, whereas if you will take just the covered box, you will end up with a million dollars. It seems a shame to sacrifice a million dollars for a thousand.

An argument for taking both boxes might run as follows: The Being has already played his move. Whatever you now do will not affect the amount of

Received 12 *April* 1972

money in the covered box. Regardless of whether the Being did or did not put a million dollars in the covered box, you stand to receive \$1,000 more by taking both boxes than by taking the covered one alone. It seems a shame not to take advantage of the fact that the Being played before you, and you play second.

Further elaborations of these arguments can be found in Nozick [1969].

Our solution is based on a game theoretical approach, following that of Nozick. Below is a representation of the game in matrix form.

		Being	
		P_1	P_2
you	A_1	\$1,000,000	\$0
	A_2	\$1,001,000	\$1,000

where: P_1—Being predicts you will take only covered box
P_2—Being predicts you will take both boxes
A_1—You take only the covered box
A_2—You take both boxes,
and the cell entries are your payoffs.

The first argument above is based on the principle of maximising subjective expected utility (*SEU*). This principle states that one should select that action which leads to the highest expected utility. The (subjective) expected utility of an action is the sum of the products of the utility of the (mutually exclusive and exhaustive) outcomes it may lead to, by their probability given this action. Thus, in the present case, the *SEU* of A_1 is $p.u(\$1{,}000{,}000)+(1-p)\cdot u(\$0)$, and of A_2 is $q\cdot u(\$1{,}001{,}000)+(1-q)\cdot u(\$1{,}000)$, where $0<p,q<1$, p is close to 1 and q is close to 0. No doubt the reader's utility function yields a higher *SEU* for A_1 than for A_2.

The second argument above is based on the dominance principle. This principle states that if for every possible state of the world (or, alternatively move of your opponent) you are at least as well off by doing one act as by doing another, and even better off for some state (or move), then you should select the former act in preference to the latter. Thus, in the present case, the possible states are that there are a million dollars in the covered box, or that there are not, and in either case you are better off by doing A_2 than by doing A_1.

This apparent discrepancy between the recommendations of the *SEU* principle and the dominance principle results from an injudicious application of the dominance principle. Consider, for instance, the following example: Israel must decide whether to withdraw from its occupied territories or not, and Egypt must decide whether to declare war on Israel or not. Suppose the following matrix represents the possible payoffs to Israel on some ordinal utility scale.

		Egypt	
		war	peace
Israel	withdraw	0	2
	remain	1	3

Clearly, remaining in the occupied territories is the dominant strategy, since $1>0$ and $3>2$. Suppose, however, that you believe that with a high probability withdrawal will be conducive to peace while remaining in the territories will eventually lead to war. Then you might prefer to withdraw and end up in the 2 cell than to remain and end up in the 1 cell.

This example makes clear that the dominance principle loses its appeal when applied to situations where the states of the world (or the opponent's moves) are affected by the decision maker's actions, and its logic is overriding only when the states are independent of the actions, *i.e.* when the probability distribution over states of the world (matrix columns) is the same for all actions (matrix rows). Let us refer to this case as the unconditional case. In the unconditional case a dominant strategy, if one exists, coincides with the maximun *SEU* strategy, and in fact with the strategy recommended by any of the other principles of 'rational' choice. This constraint on the applicability of the dominance principle, though obvious and undisputed, has nevertheless rarely been explicitly stated in the literature (see, however, Jeffrey [1965], pp. 8–10). Since in our problem the states of the world are not probabilistically independent of actions, there is no case for the dominance principle. Furthermore, it can be proven that for every finite partition of the world into states that are not probabilistically independent of actions, there is a refinement of this partition into an unconditional case. We show this for a simple example. The general case is discussed and proved in Krantz, Luce, Suppes and Tversky [1971].

Let the following be a payoff matrix:

		states	
		S_1	S_2
actions	A_1	1/4 0	3/4 4
	A_2	2/3 2	1/3 6

where A_1 and A_2 are the two possible actions; S_1 and S_2 is a partition of the world; 0, 2, 4 and 6 are the possible outcomes on some ordinal utility scale; and the fractions give the probability of the respective cells for a given action. Obviously, we do not have probabilistic independence in this case. We now show how to refine this partition to obtain another one, under which states *are* probabilistically independent of actions.

Suppose that which state of the world obtains is determined by randomly drawing chips from a poker bag. The poker bag contains twelve chips: three are red, five are green, and four are yellow. Drawing a red chip always entails S_1; drawing a yellow chip always entails S_2; but drawing a green chips entails S_2 if the action is A_1, and S_1 if the action is A_2. Let us now define S'_1 as the state defined by drawing a red chip; S'_3 as the state defined by drawing a yellow chip; and S'_2 as the state defined by drawing a green chip. This yields the following matrix:

states

	S'_1	S'_2	S'_3
A_1	1/4 0	5/12 4	1/3 4
A_2	1/4 2	5/12 2	1/3 6

The new states are probabilistically independent of your actions, yet obviously every action now leads to its possible outcomes with the same probability as before. Note, however, that A_2 is no longer the dominant strategy, since, for S'_2, A_1 is the better action. Although $[S_1, S_2]$ may be a more natural partition of the world in some sense, it does not enjoy any formal privilege over $[S'_1, S'_2, S'_3]$. Sometimes, in fact, a new partition, arrived at formally, may be interpretable in some meaningful way. Suppose, for instance, for simplicity's sake, that in our problem $p = 1 - q$, *i.e.* the probability for getting a million dollars when taking one box is equal to the probability of getting a thousand dollars when taking both boxes.

We can now take that probability to be the Being's likelihood of predicting correctly, and repartition of the world into states not merely independent of your actions, but also with a meaningful interpretation, as follows:

		Being	
		D_1	D_2
you	A_1	p \$1,000,000	$1-p$ \$0
	A_2	p \$1,000	$1-p$ \$1,001,000

where A_1 and A_2 are as above; D_1 and D_2 are the possibilities that the Being predicts right or wrong, respectively; and p is the probability of the Being predicting correctly.

The problem, when thus presented, is seen to be not really one of deciding between two principles of choice. However, doing away with the case for the dominance principle does not seem to do away with the paradoxical nature of the problem. One is left with the uneasy feeling that choosing A_1, though defensible on game-theoretical grounds, is somehow 'wrong' in a very fundamental way. That it is, in fact, tantamount to subscribing to backwards causality. In other words, choosing A_1 rather than A_2 seems essentially to be justified by the fact that thereby a very high, rather than very low, probability can be assigned to a certain desired event, namely that the Being put a million dollars into the covered box. But to go about assigning probabilities to past events, unaffected by present events, in what appears to be a completely arbitrary, *ad hoc*, and wilful fashion is, to say the least, highly unorthodox and more than a little unsettling. Before answering this objection, we would like to criticise Nozick's way of dealing with it.

Nozick persists in presenting the dilemma embodied in this situation as one of choice between two decision principles. He therefore proceeds to distinguish between probabilistic independence and logical independence, where states are logically independent of actions if the actions 'do not affect, help bring about, influence, *etc.*' (*op. cit.* p. 132) which state obtains, and then recommends as a general policy that where states are logically independent of actions, even if they are not probabilistically independent of them, 'one should perform the dominant action' (*op. cit.* p. 132).

This solution, though relieving one of any suspicion of adherence to backwards causality, poses its own threats to one's image as a rational decision maker. (*i*) It puts you in the extremely uncomfortable position of acting 'against what you would rationally want to bet on' (*op. cit.* p. 116).

(*ii*) It implies that even if you believed that choosing A_2 would *surely* mean loss of the million dollars, *i.e.* even when you believe that the Being has perfect predictive powers, you should still choose A_2 whenever the Being played his move before you, for reasons of 'logical independence'. This is a conclusion which Nozick himself is unwilling to draw, which leads him both into an inconsistency with his own maxim, and causes him to recommend a different strategy for the case when the Being's probability of correct prediction is 1 and for the case when it is $n/n+1$, for an arbitrarily large n. The distinction between logical and probabilistic independence here is at best shaky.

The question now naturally arising is whether the paradox does not lie in the very assumption of such a Being itself. In others word, the question may seem to be not what strategy a rational decision maker should employ under the proposed circumstances, but whether a rational man could ever find himself in such circumstances to begin with. What kind of evidence would lead to such overwhelming faith in the predictive powers of any Being? Can any situation lead a rational man to simultaneously believe that the Being plays his move, irrevocably, prior to your move and yet that the probability of there being a million dollars in the covered box is different if you play one strategy than if you play another? Let us examine this question first psychologically, and then logically.

Since neither the funds nor the Being for a real-life tryout of this game are available to us, we urge the reader to follow us through the following thought experiment: Suppose you have volunteered to participate in a psychological experiment at the local university. Sitting behind a one-way screen, you watch many subjects play this game against an experimenter. Time and again, you see the experimenter put into a covered box a check made out either for a million dollars (play money, of course . . .) or zero dollars; you then see a subject entering the room, receiving instructions on the nature of the game (essentially by way of its payoff matrix, and possibly with statistics on how former players have fared), and then playing his move. You note down for each subject the amount he wins in the game; to your extreme surprise you soon realise that all those who come their turn had taken only the covered box had found the check in it to be made out for a million dollars, and all those who had taken both boxes had found their check to be for zero dollars. You are now summoned to play the game yourself. The experimenter had played his move, and it is your turn. What would you do?

First and foremost, we predict, what you would do is to discredit the evidence of your senses. You would suspect foul play. You would suspect that you are being taken. You would either tell yourself that the checks were, by some clever sleight of hand, tampered with or exchanged *after* the subjects had played their move (and you need not feel any more called upon to explain how this was done than if you had been watching rabbits being pulled out of a hat); or you might believe that the subjects were actually collaborators of the experimenter, preinstructed as to how to play, *etc.*; or you would just simply think you were imagining things. Anyway, you would play your move depending on the way you interpreted the goings on (if, indeed, you were still motivated to play it rationally . . .). In all probability you would ***not***, however, be holding simultaneously to both beliefs, *i.e.* that the check was written out irrevocably prior to the subjects deciding what to play, and that the probability for subjects playing the two strategies in good faith is different (for the event of having the check written out to the sum of a million dollars). Indeed, to go back to our Being, each of the two premisses we are given would seem to be counterevidence to the other. That people who play A_1 find a different sum of money in the covered box than people who play A_2 would normally be taken as proof that the sums were put there after they made their choice; and that the money was put in the box before you choose would normally be taken as a guarantee that the probability for finding a million dollars in the box is the same whether you will now take both boxes or just one. That this is not so in our case, we claim, is maybe counterintuitive, but not logically contradictory.

Let us check our intuitions. The two seemingly incompatible premisses above are essentially equivalent to the single premiss that the Being has near perfect foreknowledge of which of two acts you will perform. Foreknowledge in itself is not a logical impossibility. If, for instance, the Being were playing not against yourself but against some robot, say, it would be perfectly plausible that he could predict how the robot would play. It does however, seem virtually impossible that *your* moves and choices are so predictable! After all, you, unlike the robot, have free will! But do you? Maybe you are only under an illusion of free will, like a robot may be programmed to be (for example, program him to make a different response than he would have otherwise every time he is told what his response will be . . .). In fact, our Being, if he existed, would contribute just the kind of evidence that would disprove your illusion that you can choose arbitrarily in the game, if you want to.

The paradox may now have been brought into clearer focus. To the extent that you refuse to take seriously the possibility of any conceivable circumstances that would lead you to believe in the existence of such a Being, the paradox for you becomes void. Prior evidence becomes immaterial to your present decision, and you would act on the 'sure thing'—

namely, take both boxes. Furthermore, even granted that such a Being could, logically speaking, exist, you might feel that its existence is such a major departure from a world in which you can trust your (tried and true) concepts of rationality, that you would not really want to test them as they are in such a world, and this again would render the paradox void. One may, however, go along and think this 'as if' situation to the limits which *are* afforded by our existing concepts of rationality. This is what we shall now proceed to do, and naturally it bears on the age old philosophical questions of what implications, if any, a deterministic point of view has to discussions of rationality, morality, personal responsibility, *etc.* In particular, in what sense can one recommend a certain course of action in a situation in which the course of action to be taken has actually been predetermined.

Suppose you are playing the following game: Someone is tossing a fair die, and after each toss you guess whether it came up on 6 or not. You receive a penny for each correct guess. Clearly, you will be maximising your expected gains by guessing not–6 on every toss. Now suppose that the die was tossed the previous day, the outcomes were noted down on a list, and you are now guessing, entry by entry, whether the number is 6 or not. Had you been able to obtain a copy of the list you could change your previous strategy to one that would ensure that you get a penny on each trial. But without such a list, you can *still* do no better than to guess not–6 on each trial! The fact that the order of 6's and not-6's is predetermined, given that you do not know what it is, does not affect your strategy. What this example serves to point out is that the mere knowledge that things are not what they seem to be does not necessarily supply you with an alternative strategy for dealing with them.

Thus, in our case, although the facts really imply that there is no free choice, the illusion of free choice persists, and you can do no better than to behave as if you do have free choice, *i.e.* 'deliberately' pick that strategy that seems to serve your interests best. The fatalistic claim that in a deterministic world it makes no difference which mode of behaviour one chooses because '*que sera sera*' is obviously false on statistical and phenomenological grounds, if no other. Furthermore, to follow the same argument, though you know that the Being plays before you, you nevertheless cannot do better than to play as if he plays after you. For you cannot outwit the Being except by knowing what he predicted, but you cannot know, or even meaningfully guess, at what he predicted before actually making your final choice. This feature, strange though it may seem, is built into the story by virtue of investing the Being with such astounding predictive powers, and of *behaviour*, no less, rather than just inclinations. (As a side remark, let us say that though we do not wish here to embark on speculations as to how such a Being might go about arriving at his predictions, it seems safe to say that it is *not* by a process that you could ever hope to master yourself. Since if you did, like the Being, know what you were going to do before doing it, you could intervene in the course of events, so to speak, and disprove that prediction by in fact behaving otherwise.)

Finally, note that the uneasiness created by the persistent feeling that choosing A_1 over A_2 amounts to arbitrarily choosing to assign a high rather than a low probability to a past event is shown by this analysis to be no more than an illusion itself; for if indeed your feeling that you are a free agent and can choose arbitrarily between the strategies is illusory, so, of course, is the feeling that probabilities are here being assigned in an arbitrary and *ad hoc* manner.

To sum up, we have tried to show that the Newcomb problem allows for just one, rather than two, 'rational' strategies (given that what you are trying to do is get as much money as you reasonably can. One could, of course, advocate the taking of both boxes on maximin grounds; *i.e.* assuring yourself of at least a thousand dollars.) This strategy is to take only the covered box. It is not justified by arguing that it *makes* the million dollars more likely to be in that box, although that is the way it appears to be, but because it is inductively known to correlate remarkably with the existence of this sum in the box, and though we do not assume a causal relationship, there is no better alternative strategy than to behave as if the relationship was, in fact, causal. The pragmatic nature of this argument is supplemented by other considerations which serve to show that any logical inconsistencies which seem to be attached to choosing A_1 rather than A_2 are due to the unintuitive implications of the existence of such a Being, but are not real contradictions, merely illusory ones.

Thus, though we began by analysing the problems along the lines set out by Nozick, we ended up with the opposite recommendation to his. Our solution manages to avoid the inconsistencies attached to Nozick's solution, but nevertheless, we hope, succeeds in meeting the objections that drove Nozick away from offering it himself. By proposing a rationalisation for a strategy that seemed to be a mere act of faith, we hope to convince the reader to take just the one covered box, and join the millionaires's club!

The Hebrew University of
Jerusalem

REFERENCES

Jeffrey, R. C. [1965]: *The Logic of Decision.*

Krantz, D. T., Luce, R. D., Suppes, P., Tversky, A. [1971]: *Foundations of Measurement, Volume 1.*

Nozick, R. [1969]: 'Newcomb's Problem and Two Principles of Choice,' in N. Rescher (*ed.*)*: Essays in Honour of Carl G. Hempel*, pp. 114–46.

Brit. J. Phil. Sci. 23 (1972), 305–327 *Printed in Great Britain*

Biological Classification*

by VERNON PRATT

Introduction.

The intuitively recognised distinction between natural and artificial classifications.

1 *Some questions distinguished.*

What is the classification's basis? Examples of answers:
- (a) evolutionary relationships
- (b) character of the reproductive apparatus
- (c) character of the genotype

What job or jobs is the classification expected to help with?

(Gilmour's suggestion that the naturalness of a classification hangs on the number of functions it performs)

How is it that using a particular basis will produce a classification which will serve its intended functions?

How are taxonomically important characters to be picked out from those that are unimportant?

2 *Some Theories of biological classification.*

- (*a*) *Archetype Theory.*
  Cuvier's two principles of weighting.
- (*b*) *Phylogenetic Theory.*
  Genotypical elements in at least one important exposition.
- (*c*) '*Overall resemblance*' *theories.*
  Criticism of phylogenetic views.
  Cain's claim that the naturalness of a classification depends on the number of characters taken into consideration.
  Two ways of 'taking every character into consideration':
    - (*a*) classification on the basis of numbers of shared characters. This gives the numerical approach.
    - (*b*) classification on the basis of *importance* of shared characters.

Received 6 *September*, 1971

---

* I wish to thank Professor Peter Alexander, Professor Dorothy Emmet, Professor D. L. Hull, Dr Humphrey Palmer, Dr Roger Woolhouse, the Editor and the members of the L.S.E. Philosophy Seminar for their comments on an earlier draft of this paper, and to acknowledge the hospitality of the University of Ife, Nigeria, where it was written.

The latter does not represent an independent theory at all. The former is incoherent, for there is no such number as the number of characters which an organism has.

Genotypical element in numerical taxonomy.

3 *Towards an improved theory.*

The possibility of induction a precondition of science. Setting up inductive groups the object of 'natural' classification.

Species are natural groups because the individuals they unite have similar genotypes—similar in the sense that each of their component instructions is interchangeable with all the others it corresponds to.

Higher taxa may be natural groups if they unite individuals which have the same organisation above a particular level.

Evolution has in fact generated such groups, so that a phylogenetic classification of the right sort will be natural.

Conclusion.

INTRODUCTION

From time to time biologists profess themselves to be searching for what they call a 'Natural Classification', and in what follows I want to propose an answer to the question of what it is they are looking for.

It may be helpful at the outset to indicate intuitively the sort of notion of which we shall be seeking a rigorous or philosophical understanding.

If we take a group of objects, say the plants growing in a garden, we can divide them into groups in innumerable different ways. We could, for example, classify them according to whether or not their stems were over a foot in length. Or we could group them according to their smell, putting plants with attractive smells in one group and those with unpleasant smells or none at all in a second. Or we might set up just twelve categories, and ascribe individuals to them on the basis of their month of flowering.

It might be said, however, that in following any of these procedures we would often be involved in separating individuals which on the basis of untutored observation seemed very similar. On the first criterion, for example, what according to orthodox modern taxonomy we should call two individuals belonging to the same species (say, the White Dead-Nettle) might be allotted to different groups since only their height is taken into consideration; and observation, even unprejudiced by orthodox taxonomy, would suggest strongly that two such individuals were very like each other, in spite of the difference in height. When the same unprejudiced eye is cast over the rest of the individuals belonging to the two groups thus created, it is likely to report that not only are individuals apparently very similar to each other being separated, but some individuals which seem on every count except height to differ radically from each other are being grouped together: the apple tree goes along with the rambling rose, the Shepherd's Purse with the Sphagnum moss (using for the sake of reference the categories of orthodox taxonomy).

Both of the other two criteria we suggested would give rise to the same feeling: the feeling perhaps we might say of *arbitrariness.* In using these criteria we would seem to be grouping the individuals without any real regard for the resemblances, relations, and differences which exist among the plants themselves: we are, it could be said, classifying them *artificially.* Such an expression holds out the hope that it might be possible, as an alternative, to classify these individuals (and organisms in general) non-artificially or *naturally*. Taxonomists (botanical, in the following quotation) would then be looking in the words of William Whewell, for something 'not of their own creation;—not anything merely conventional or systematic; but something which they conceive to exist in the relations of the plants themselves;—something which is without the mind, not within;—in nature, not in art;—in short a natural order' (Whewell [1840], volume 1, p. 474). Pierre Duhem expresses the same intuition, this time as regards the zoologist. In looking for a natural classification he would be looking, Duhem says, for a system whose relations 'correspond to *real* relations among the associated creatures brought together and embodied in his abstractions.' (Duhem [1906], p. 25, my italics).[1]

So much, then, by way of indicating (simply by an appeal to intuition) that there is a distinction between artificial and natural classification. But is our intuition correct? Is there such a thing as a natural as opposed to an artificial way of classifying organisms? If there is, what are the grounds upon which the distinction is made?

1 SOME QUESTIONS DISTINGUISHED

Several questions arise, connected though distinct; and an understanding of the theorising about classification that has gone on (largely in prolegomena to works devoted to substantive taxonomy) depends in large part on distinguishing between them. Most classifications, and all the biological classifications that have seriously been canvassed, are based on resemblances and differences between things. Things that resemble each other are placed together, and separated from other things from which they differ. But this is an almost completely formal statement, because the notion of two things *resembling* each other remains empty until the respect or respects of resemblance are indicated. With dubious exceptions the members of any pair of things resemble, and differ from, each other in an indefinite number of ways.[1] 'Resemblance' as Renford Bambrough says, 'is unintelligible except as resemblance *in a respect.*' (Bambrough [1961], p. 204).[2] In defining a scheme of classification therefore, whether a new one is being proposed or an attempted made to provide a rationale for an existing one, we must have an answer to the question: What are the respects (or what is the respect) in which we are to compare things in order to apply to them the classification in question? We might express this as the question, What is the classification's basis?—or, What is the principle of unity embodied in its groupings?

---

[1] Duhem is interested in using the notion of a natural classification to illuminate that of a scientific theory, rather than in naturalness itself. His account of the latter is completely undeveloped: to say that a natural classification is one that 'corresponds to reality' is of course to *pose* the philosophical problem, not to answer it.

One recommendation that has often been made is that we should look to the evolutionary relationships that hold between organisms and classify them on that basis. Thus if two species are alike in springing from a certain ancestor-group, and a third originates from a different source, that might be a reason for classifying the first two together and separately from the third. This would be, however, just one possibility, for there is more than one way of classifying organisms phylogenetically and different people have recommended different ways (*below*, §2(b)). For a full statement of the respects in which organisms were to be compared we should have to go into each of the varieties. Nevertheless, 'evolutionary relationships' constitutes a distinctive outline answer to the question (let us call it the first) we have just distinguished. 'In respect of their reproductive apparatus' is a second, which has also been widely influential in the past; and 'In respect of their genotypes' a third, for obvious reasons of more recent importance.

I stress that all these answers give an outline indication only, and we shall explore two of them in a little more detail later (section 2). But at present we are simply concerned to distinguish as sharply as possible between the different questions which a complete rationale of a classification must answer.

A second question asks what job or jobs a classification of living things is expected to help with. Some have thought important the rather humble job of aiding the memory (*e.g.* E. Pigeon on his preface to Cuvier's *Le*

---

[1] For this reason the following definition of a natural kind seems less than illuminating: 'A Natural Kind is a group of objects which have *many* (perhaps indefinitely many) features in common' (Price [1953], p. 7).

[2] Or Popper: 'Two things which are similar are always similar *in certain respects*' (Popper [1955], p. 420; his italics).

Régne Animal (Cuvier [1817])), and something of the same idea may be at the bottom of the more modern-sounding suggestion that classification should above all provide an instrument for storing biological information in such a way as to make it easily accessible (Mayr, [1969], p. 78; Heslop-Harrison [1962], p. 14.) Another closely related idea is that it should enable us to express our biological knowledge in the most economical way possible. Cuvier says, for example, that a soundly based system of classification will provide the means of expressing the properties of the individuals so classified in the fewest, most comprehensive, most significant terms. Mill thought a classification ought to serve the heuristic function of generating hypotheses (Mill [1843], book 4, chapter 7); Mayr is one who defends the further idea that the hypotheses generated should have a good likelihood of being true, so that one of the functions of a classification on his view is that of generating reliable predications (Mayr [1969], p. 79–80. See also Hempel [1965], p. 146.)

There is no *a priori* reason why the performance of one of these jobs should preclude the performance of any of the others; and one might easily think that the more jobs a classification does, the better it is. Indeed, the philosopher J. S. L. Gilmour, for example ([1951], p. 401), has suggested that it is the capacity of a classification to serve a number of different functions which makes it 'natural' as opposed to artificial—although his suggestion is I think misconceived since it implies that functions can be counted independently of being specified (cf. the parallel point about the notion of a character, *below*, §2(c))[1]; as I hope to make plain as we proceed we are likely to achieve a better understanding of the natural/artificial distinction if we wonder *why* some classifications are more generally useful than others.

The two questions so far distinguished clearly bear on each other. It might well seem that the logical procedure (if one were to begin classification *ab initio*) would be to decide what functions the proposed classification would be required to fulfil, and then to choose the principle of unity likely to create the classification best suited to those ends. But such a possibility immediately points to a third question: How is it that employing a particular principle of unity leads to a classification which serves its intended functions? For example, Why should it be that a phylogenetic basis for one's classification produced the most efficient device possible for storing retrievable biological information? (Mayr[1969], p. 78.) Or, What

[1] See Simpson's comment in his [1961], p. 25, and Heslop-Harrison's [1962], pp. 21–2. Gilmour makes a different suggestion also, that 'there will be one, as it were, general-purpose classification, more natural than others, in the sense that more inductive generalisations can be made about its classes' (Gilmour [1951], p. 402). This and other remarks in that paper are a good deal closer to the view I try to defend in the present paper.

is the mechanism which secures that groups set up on a genotypical basis are helpful in deriving reliable predictions?

Depending on the answers given to the other questions this third one may or may not arise. We do not, for example, need to have an explanation of how a classification based on phylogeny provides us with a means of identification: any classification worth the name would surely do that. But in other cases it is not obvious why the suggested basis should yield a classification capable of doing the job assigned to it, as in the two examples just mentioned.

Also, the kind of answer required by our third question may differ from case to case. Sometimes, what needs to be shown is a logical connection between the principle of unity and the capacity of a classification to perform a certain function. Some theorists for example have urged that an important task of classification is the provision of a framework which will allow biological knowledge to be expressed 'economically'—and the greater the economy, the better the classification. When some of these theorists also suggest that 'overall resemblance' should be the basis of classification (see *below*, §2(c)) it may be that they have in mind a connection of this logical kind between economy of expression and the basis they suggest. On the other hand, our third question often demands factual information. It relating a phylogenetically based classification to its ability to perform the function of generating reliable predictions, for example, reference will very likely be made to our knowledge of evolution and genetics.

The practising taxonomist will however be interested primarily in a question that has still to be raised: How are taxonomically important characters to be picked out from those that are unimportant? If we are trying to group organisms phylogenetically, for example, which of their characteristics can we regard as indicative of evolutionary relationships, and which can we ignore as unhelpful? For the practical task of classifying we shall need to know some principles whereby this distinction may be made.

Once again, it seems (at first sight anyway) that for some principles of unity this fourth question does not arise. It has been held, for example, that in grouping things according to 'overall similarity' it is possible, and greatly desirable, not only to take every character into account, but to regard each as having the same taxonomic importance as any of the others. If such an idea were a coherent one (though I shall argue that it is not) it clearly rules any criteria of taxonomic importance out of court. But equally clearly for some principles of unity, our fourth question is legitimate and must be answered.

2 SOME THEORIES OF BIOLOGICAL CLASSIFICATION

We turn now to consider three (and a half) 'theories' of biological classification as a preliminary to presenting our own. None of these theories is silly, and they all—even the theory of archetypes which we are to discuss first and which is often represented as a pre-scientific phantasy—contribute substantially to the fuller understanding that now, I believe, is possible. In many ways that fuller understanding comes of seeing how these several insights can be synthesised into a single account.

(*a*) *The Archetype Theory.*

It is assumed by the theory of archetypes that we have to regard the living organism as constructed, or derived from ancestors which were constructed, according to a 'blueprint' or plan—or, to use its own word, and 'archetype'; and it holds, accordingly, that we ought, in classifying organisms, to form our groups according the the similarities the blueprints associated with the organisms bear to each other.

To begin with, the taxonomist will form a series of groups each consisting of organisms 'built' according to the very same plan as each other: we may think of this as constituting classification at the 'species' level. But he will find that some plans, while not being identical in detail with certain others, are yet identical with them at a certain level of generality. So arises the possibility of bringing together the first order groups into higher level categories, where the plans associated with each member group share a common general outline. Parallel considerations lead us to admit the possibility of forming third order groups, and fourth, and so on: the groups of a given order uniting some of the next lowest rank in virtue of the fact that their plans, at a certain level of generality, possess a common form. Thus would be generated the hierarchy of categories we are familiar with in orthodox classification—individuals being united into species, species into genera, genera into families, and so on.

How, on the archetype approach, is the practising taxonomist to know which characters to regard as taxonomically significant, as reliable indicators of archetype relationships? The answer of one such theorist, Cuvier, is instructive. First, he suggests, we should regard as important the characters which heavily influence the rest of the organism—those which are 'most essential' to its life-form. As Cuvier himself puts it: 'The parts, properties, or traits of conformation which . . . exert the most marked influence over the ensemble of the entity, are those which are called the ***important characters***, the ***dominant characters***' (cp. Cain, [1959], p. 188).

This suggestion clearly stems from the emphasis Cuvier places throughout his zoological work on regarding the organism as a 'whole' made up of complexly interacting parts. You knew, Cuvier thought, that if an organism possessed a certain kind of respiratory system, for example, the rest of its internal organisation must fit in with this: the various component parts which together made up the whole system which is the organism must be correlated with each other. Some features would clearly have wider implications than others: the colour of an organism would not have theprofound effect on its total organisation as would say, its circulatory system. It is those with the larger implications which Cuvier, reasonably enough, is suggesting are the best guide to what perhaps we may call, rather tendentiously, an organism's 'fundamental organisation'—its archetype.

Cuvier's second criterion is less interesting. We should consider those characters as most important, he says, which are most constant throughout the group, so that the circulatory system must be regarded as significant when we are considering mammals, since although other features vary from mammal to mammal, the circulatory system remains the same. But as Cain argues ([1959]), when directed at taxonomists attempting to set up groups of organisms, this second suggestion is patently circular. If our project is the selection of characters to use in establishing groups, we cannot appeal to *constancy*—making reference as it does to groups which *ex hypothesi* have not yet been set up.

Cain argues that Cuvier's first criterion is circular, too. Most of the physiological theory which provided the means of assessing the influence a particular feature exercised on the functioning of the whole was itself based, Cain tells us, on observing which features were constant within the already recognised groups. Be that as it may, modern physiology is differently founded, and I believe an adequate theory of classification must take Cuvier's first criterion very seriously. The insight on which it is based had been vindicated by subsequent discoveries, particularly in genetics, and forms an essential element of the view we shall shortly develop.

(*b*) *The Phylogenetic Theory.*

Those who take the evolutionary or phylogenetic view argue that our grouping of organisms should reflect the attempt to set forth the groups and relationships between groups which have been produced by the process of evolution. This general programme can however be realised in a number of different ways: it is in fact inadequate to speak of *the* groups produced by evolution, for at least at levels higher than the species the workings of evolution can be thought of as producing different groupings depending on which aspect of those workings is stressed.

If you have, for example, three groups, *A*, *B*, and *C*, two different questions about their evolutionary relationships arise. The first concerns how long ago and by what route they each arose from their common ancestor-group. If we classify simply on the basis of this consideration we might put A and B together and separated from C on the grounds that the ancestral lineage gave off C a long time ago, and then subsequently split, to form A and B.

A second question arises, however. Because the rate of evolutionary change can vary widely in different lines, two groups with a common origin may be genetically and phenetically either closely similar or widely dissimilar, depending on the degree of divergence that has occurred. If we suppose B in our previous example to have diverged rapidly from A, which, let us suppose, has evolved in parallel with C, there would be a case for classing A and C together and apart from B.

The first scheme, which only takes into account the order in which the groups in question diverged from a common ancestor-group, is most conveniently called 'cladism' (Mayr [1969], pp. 70 ff.); and its results will be quite different from those which ensue if you take the relative divergence of groups into the reckoning.

The important question thus arises: What criterion or criteria can be appealed to in order to choose between these alternative schemes? Our own remarks would suggest that we should choose in the light of the job or jobs we want our classification to assist us with. Ernst Mayr, as a defender of the second alternative against the 'cladism' of W. Henning and his followers, seems to adopt this approach. He argues that the ends of classification will be better served if the groups one sets up are a true reflection of the relationships that exist among the genetic programmes, or genotypes, of the organisms concerned: 'When we classify organisms, classification by phenotype is only the first step. As the second step we attempt to infer the genotype, the evolved genetic programme, which has a far greater explanatory and predictive value than the phenotype' (Mayr [1969], p. 77).

But this crucial reference to the genotype—and it is made throughout his book—makes one ask whether Mayr is a true 'phylogeneticist' in classificatory theory at all. It shows that it is not fully revealing to say that for Mayr the principle of unity in classification is simply evolutionary relationships. He wishes rather to erect groupings on the basis of evolutionary relationships *in so far as these are correlated with resemblances and differences between genotypes*. As it happens, evolution has produced groupings of genetically similar organisms; but if it hadn't, we would in classifying ignore it and concentrate on genetic relationships directly. As Mayr himself puts it: 'Evolutionism, as a philosophical basis for classification, is a valid approach only if and when natural groups of organisms are the result of divergent evolution. If reticulate evolution were common owing to the frequent fusion of previously separated evolutionary lines, or if convergence were frequently so complete as to lead to groupings that could not be unmasked as having a polyphyletic origin, then the claims of evolutionism as a proper theoretical basis for classification would indeed be questionable' (Mayr [1969], p. 78–79).

We could put this equally well I suggest by saying that evolutionary relationships are not, for Mayr, the ultimate principle of unity which biological classification should employ: rather he is suggesting it should classify according to genetic similarities, using evolutionary relationships as simply an extremely useful guide. It is when we begin to answer our fourth question, about the characters we are to take as taxonomically significant if our object is to classify according to genetic similarity, that evolution comes in. Because by and large evolution has produced groups of genetically similar individuals, knowing in some degree how evolution works will assist us in discerning the aspects of the phenotype—that is to say, roughly, the observable characters of an organism—which serve as the most reliable indicators of the genotype.

The 'phylogenetic' view therefore as we see it developed in one of its most powerful advocates is not quite what the term might suggest on first acquaintance. And in going on now to consider a different kind of approach to classification theory that has been canvassed in recent years we shall find there, most interestingly, an exactly parallel state of affairs: a theory which despite every first appearance makes an appeal to genotypical considerations for its ultimate rationale.

(*c*) *'Overall Resemblance' Theories.*

Critics of the idea that evolutionary relationships should form the basis of classification have usually urged, as their principal objection, that we possess too little information to enable us to classify organisms in this way. Partly because our factual knowledge is at present so incomplete, but partly also because of the phenomena of homoplasy (convergent evolution, parallelism, mimicry *etc.*—see Simpson [1961], p. 78) it is all too frequently impossible to discern the precise evolutionary relationship that obtains between organisms. To recommend basing classification on phylogeny is therefore, it is claimed, to suggest founding it upon ignorance. (Sokal and Sneath [1963], §2. 3; Cain & Harrison [1960]; Thompson [1952], *etc.*) So the critics argue. Instead it is proposed that classification should be based on 'the resemblances existing *now* in the material at hand' (Sokal and Sneath [1963], p. 55), that is, on 'a general comparison' between organisms which takes into account 'every phenotypical character'. We are to compare organisms as a whole with each other and classify them on the basis of the 'overall similarity' we will thus discern.

Some theorists who recommend this view see it as contrasting favourably with the principles of some classifications in the past which, based on comparisons involving just one or a few characters only, were for that reason to be regarded as 'artificial'. A. J. Cain declares, in a chapter not entirely consistent with itself, that the only difference between a natural and an artificial system is that more characters are taken into account by the former than by the latter: 'Every possible intermediate between an obviously artificial and an obviously natural classification can exist. All depends on the number of characters taken into account' (Cain [1954], p. 18).

We may notice the strangeness of Cain's view here. His notion of 'natural' does not I think square with other people's idea of it, either historically or in our own time; not does it mark a distinction which promises to be of much use. Surely much more important for the nature and value of a classification than the number of characters that have to be taken into consideration in applying it is the criterion or criteria by reference to which the 'consideration' is conducted. I might, to take an instance, try to classify plants into those that would be suitable to give as gifts for Christmas and those which would be inappropriate for this purpose; and in asking my question Is this plant suitable? I might try and take every feature into consideration. Cain would have to call this a natural classification; but, according to *my* reading of biologists' discussions (I must simply leave this for others to judge—a list of references would have to be too long to prove my point) I don't think many people would agree with him. And would there be much point in pressing the word 'natural' to this unaccustomed use? I cannot myself see the value of making the distinction Cain wants the word to mark.

Cain's view of ' naturalness' (as expressed in the quoted sentences) is not alone in feeling the prick of the point I raise here however: it is seriously damaging to the concept of classification on the basis of 'overall similarity' in its entirety.

Let us try to imagine a taxonomist attempting to classify according to 'overall similarity'. He will begin, we may think, by comparing each individual with each other individual in turn and drawing up a table showing the characters each such pair of individuals shares. But once this initial task is completed (our imagination has to transcend the limitations of logic here, as I shall argue later) he must go on to interpret his findings. How are his data to be processed in order to result in the required classification?

Two possibilities seem open. The first, and one might think the most simple-minded, is to add up the total number of similarities between the members of the pairings and then to classify on the basis of numbers of shared characters. Individuals with many characteristics in common would thus be classified together and away from individuals with which they shared few common characters. We may call this the numerical approach.[1]

Contrasted with it is the probably more intuitively appealing alternative of regarding some characters as more important, taxonomically, than others. Individuals would then be grouped together if they resembled each other in *important* ways, and away from those with which they had little *important* in common. Those who take this second alternative however must have an answer to the question: By what criterion or criteria is the importance of a character to be judged? For in rejecting the notion that all characters are to count equally they must be appealing, tacitly or explicitly, to some principle whereby differential weight is to be attached to them.

At this point, this second alternative of the proposal to classify on the basis of 'overall similarity' collapses into a type of proposal we have already considered. Cain, to revert to the chapter we referred to above, suggests that characters are taxonomically important insofar as they reflect the 'basic theme' of the organism, or its 'general plan' (Cain [1954], p. 19), which makes him (at least in the chapter in question), a contemporary exponent of an archetype theory.[2] Others have held that the taxonomic importance of a character hangs on whether it offers a reliable indication of evolutionary relationships: clearly this is the phylogenetic view already discussed. And still others have proposed that it is its power to indicate the genotype of the organism that confers taxonomic significance upon a character.

Thus the non-numerical version of the 'overall similarity' view is not to be distinguished as an independent theory at all. It quite fails to provide a distinctive answer to the first of our analytic questions (What is the classification's basis?): the true basis of a proposed classification, be it phylogenetic, or genotypical, or archetypal, or something else, will only be indicated by citing the principle of weighting to be employed.

Is there anything better to be said for the numerical version of the 'overall resemblance' approach? Although it has enjoyed a certain amount of intelligent and enthusiastic support in recent years it seems to me that this doctrine too suffers from a fundamental weakness of principle. The taxonomist is recommended to classify on the basis of judgments of the form 'Individual A has more characters in common with B than it has with C' and the possibility of judgments of this form entails that an organism can be regarded as having a fixed and definite number of characters. ***But there is no such number as the number of characters that an organism has.*** Any individual living thing can be described in a variety of ways, and the number of characters it is thought of as possessing will depend on the description being employed. Of course, once a description has been given, characters can be counted: but the number is 'description-relative'. To speak as though a given individual had a fixed and definite number of characters is thus like speaking of a bag of flour as though it contained a fixed and definite number of heaps: in fact, you can only counts heaps once the flour has been poured out, just as you can only count characters once the individual has been described in a particular way.[1]

---

[1] Sophisticated statistical techniques are applied in order to erect a hierarchy of categories, but this leaves the principle of the method untouched.

[2] Cain's remarks on p. 188 of his [1959] are consonant with this position, but in later papers (*e.g.* his [1962]) he seems to be advocating a 'numerical' position. 'An analysis of the taxonomist's judgement of affinity', over which he collaborated with G. A. Harrison, did much indeed to initiate this approach (Cain and Harrison [1958]).

Numerical taxonomy therefore does not in my view present a coherent theory of classification (and I have provided elsewhere the very necessary expansion of the argument I sketch here—see my [1972],): but its interest does not stop at this point. For if the writings of its proponents—we consider specifically here *The Principles of Numerical Taxonomy*, by R. R. Sokal and P. H. A. Sneath ([1963])—are examined carefully, some curious elements are to be discerned, quite alien to their surroundings, which have somehow wormed their way into the numerical taxonomist's arguments.

In the particular case of the authors we have mentioned the plain fact is that while thay declare in some places (as we have seen) that their proposal is to banish all differential weighting, in others they explicitly embrace a principle of weighting: in some places they urge that all characters are to be considered, and all equally, but in others they discuss how taxonomically significant characters are to be picked out from those that may be safely ignored. Here we are concerned not so much with the contradictions which these writers get into but with the principle of weighting which in spite of themselves, it seems, they feel it right to employ: for, very interestingly, it comes close to the genotypical principle we found defended by Mayr.

To be used for taxonomic purposes, think Sokal and Sneath, characters should reflect the genotype of the organism concerned: 'It is generally considered that only genetically determined characters should be used in orthodox taxonomy, and with this we concur' (Sokal and Sneath [1963], p. 92). Again: 'It is undesirable to use attributes which are not a reflection of the genotypes of the organisms themselves' (Sokal and Sneath [1963], p. 66). I think it emerges clearly from their discussion that despite their own self-interpretation they in fact look to genotypical similarities for their principle of unity.

---

[1] Compare Gilmour [1961], p. 100.

The importance of the hints we find in Mayr, and then echoed in the ostensibly opposed writings of Sokal and Sneath, will become clear as we pass on now to make a more positive approach to the problems we have seen others trying to solve.

3 TOWARDS AN IMPROVED THEORY

I begin with the conviction that the significance of the job which a classification is required to perform has been underestimated by many taxonomic theorists and grossly underestimated by the rest.

Unless the biologist is able to think of the individual organism he is studying as a representative of a group his results cannot be anything other than particular statements, asserting truths about individuals. The limitation of such a predicament will be obvious. For science to be possible at all, individuals must be referrable to categories, and the categories must be such that a scientist must be able to think that when certain circumstances obtain in establishing facts about particular individuals he is thereby establishing facts about members of the category in general.[1]

Inferences of this general kind must have a place in every science[2]—their justification of course constitutes the celebrated problem of induction—and to recognise the central role of induction in biology in particular—is to recognise the central importance of assigning organisms to groups such that induction is, in certain circumstances at least, reliable within them.

While it may be right therefore to say that there is no limit to the number of possible classifications of biological objects (Bambrough [1961], p. 203, to take a recent example) it seems to me to be wrong to deny (as Bambrough does) that any may be more fundamental than others. A classification of biological individuals into groups within which induction works seems more fundamental than any other, in the sense that without it no scientific biology could go on. This suggests I believe the most satisfactory way of construing the dog-eared question of whether there are 'natural kinds' of organisms: it is best understood as the question of whether organisms *can* be sorted into groups of this fundamental kind. And a classification into such groups has best title, I suggest, to the accolade of 'natural'.

---

[1] This should not be taken to imply that I believe his findings will be best represented by expressions of the form 'All *A*'s are *B*'s: it may be true that the Herring-gull has a red spot on the beak and yet false that all Herring-gulls have red spots on their beaks (some might have lost their beaks through disease, for example). But in inferring truths about 'the' Herring-gull from truths about the particular Herring-gull or gulls actually studied the scientist is nonetheless involved in making claims about individuals other than those actually studied.

[2] But the details differ. In chemistry, for example, a study not of things but of the substances of which things are made, the inferences are from a truth about a specimen of the substance to a truth about the substance in general. This brings out one reason why it is misleading to speak of 'the' problem of induction.

If this approach is right, our object in biological classification is to set up what we may call for short 'inductive groups'. There is of course no guarantee *a priori* that such an object is achievable. One can easily imagine groups of things within which inductive inferences are not in any way reliable, and there would be no way of telling, in advance of acquiring relevant factual information, whether biological organisms could be grouped more 'naturally' than this or not. The success of biology as a science, on the other hand, however, shows that as a matter of fact its inductive inferences have often been sound—the literature is full of results which later experience has substantiated—and so we have it confirmed that much, at least, of the classification actually employed in biology is natural.[1] Largely by intuition—by the method of *tâtonnement* or 'groping in the dark' of which de Candolle speaks (see Cain, [1959], pp. 202–3)—biologists have been able to set up groups within which some measure of reliability in inductive inference is achieved.[2]

Our problem is therefore that of explaining how this has been possible. Why have biologists been able to set up natural groups? What facts about living things gives the search for a natural classification the hope of success? In applying ourselves to this question we shall hope for illumination on another score, too. The practising taxonomist will need to know how to translate the very abstract injunction to set up inductive groups into practical terms. How are such groups to be erected? By what empirical criteria shall we be able to tell whether *A* and *B* are to belong to the same group but *C* and *D* to a different one?

Our strategy in what follows will be to show how a sophisticated phylogenetic-genotypical theory such as Mayr's comprises a set of principles which, generally, have the effect of setting up the type of groups we have specified. In setting up groups which reflect (in a certain way) evolutionary relationships, we are, it will be argued, setting up inductive groups. For an answer to the practical question mentioned above we can therefore refer to such texts as Mayr's and Simpson's where the empirical criteria for setting up phylogenetically based groupings are discussed.

Why is it then that the appropriate kind of phylogenetically based classification sorts organisms into inductive groups? In outline the answer is clear, and we have already noticed Mayr indicating it. The sort of phylogenetic grouping he recommends sorts individuals into groups whose members have similar genetic endowments; and because the form an organism takes in a given environment is determined by its genotype, similar organisms (in similar environments) will result from similar genotypes. If a fact about an individual organism is established therefore, it seems that we can reliably infer that similar things will be true of all individuals (in the same environment) which have similar genotypes. In other words, groups of organisms with similar genotypes which will result from the appropriate kind of phylogenetic classification, will be inductive groups: and thus the required explanation is provided.

---

[1] 'That there are or have been regularities, for whatever reason, is an established fact of science' (Quine [1969], p. 13).

[2] I say 'groping in the dark' because inductive groups were being set up before the genetic facts which provide their rationale were known. Hull reads something more than this into '*tâtonnement*' (Hull [1967], p. 180). But this is a point of interpretation. On the point of substance see Quine [1969], p. 15.

Yet it was right to say (above) that we have here an *outline* explanation only, for the import of the convenient word 'similar', very busy in the account just presented, has not been sufficiently indicated. As we have already pointed out, with a few dubious exceptions any one thing is similar in some respects to any other, and in particular any two genotypes resemble each other in some respects as well as differing from each other in others.

If, therefore, one wishes to claim that the species (say) is a natural category in virtue of the fact that individuals belonging to the same species have similar genotypes, one must go on to specify the similarity. It is to this task that we now address ourselves.

It is important to realise at the outset that not all taxonomic categories in orthodox use need have the same rationale. The facts that make a category natural need not be the same in every case. Actually we shall see that a species—'biologically' conceived as an interbreeding natural population reproductively isolated from other such groups (Mayr, [1969], p. 26)—is a natural group for reasons other than those which make higher groupings natural. Accordingly, our exposition treats of the two types of group separately.

How then are we to specify the similarity that holds between all genotypes belonging to the same species? In the following way: two genotypes from the same species resemble each other to the extent that in general any gene of one could replace the corresponding gene of the other and form a genotype which would produce a viable organism. Corresponding genes (alleles) within the species can thus generally be said to be 'interchangeable': the system to which a genotype gives rise (*i.e.* the organism as a whole) is able to function successfully irrespective of the particular allele which occupies any given locus.[1]

---

[1] It is true that the gene pool of a species can contain *some* genes which in certain combinations produce non-viable offspring—*e.g.* the homozygous occurrence of the gene responsible for sickle-celled anaemia. But these, I think, must be exceptions to the general rule.

What we assert here is the consequence of certain genetic facts, which may be briefly rehearsed: an organism's genetic complement comprises a set of chromosomes derived from one parent, together with a set of chromosomes derived from the other parent, each member of which corresponds with just one chromosome (its homologue) in the other set. During the generation of gametes homologous chromosomes pair and genetic material is interchanged between them in a process called 'crossing-over'. Generally, crossing-over can occur at any point along the paired chromosomes, and so *any* element of one can be exchanged for its corresponding element in the other. In time, therefore, it is possible for any gene to be replaced by any of its alleles: and this means that if a high proportion of viable genotypes (*i.e.* genotypes capable of generating viable offspring) are to be produced corresponding genes must be in general intersubstitutable: a viable genotype must result no matter which of a gene's alleles occupies a given locus.

A species is in this respect like a fleet of cars, the corresponding parts of which are interchangeable. A manufacturer, briefed to hand-build such a fleet, would have to build each so that it is closely similar to each of the others: and the exact sense of 'closely similar' here is given by the indicated requirement, *viz.* that each car must be such that its functioning would not be seriously impaired by the substitution for any of its parts of a corresponding part belonging to any of the other cars of the fleet. Minor variations among the cars would of course be possible. But they would have to be so minor that they had no bearing on the working of the wholes to which they belonged.

It is the fact that the genotypes belonging to a single species are composed of 'interchangeable (corresponding) parts' that makes the species what we have called an inductive group. For it implies that there is a level of discussion at which in talking about an individual one is also talking about all the individuals of the species. There is also of course a level at which variation is possible within the species—a gene's different alleles each have different effects. But very few of those effects must be of such a kind as to prevent the successful functioning (in any of its possible combinations with other genes at the other loci) of the resultant organism; for otherwise too many non-viable offspring would be produced for the species as a whole to compete successfully with its rivals.

Suppose then that one discovered a certain fact about an individual member of a species. It might be a fact which we could reliably infer to hold of all the other members of the species; but it might be a fact peculiar to the particular individual studied. How are we to decide between these two possibilities?

The answer suggested by our remarks above is that we can tell by bringing to bear, after first acquiring, an understanding of the workings of the functioning system which is the organism in question. If the discovered fact is of central importance to the functioning of the system as a whole then it can reliably be regarded as holding generally; but if it has little bearing on the system's working to regard it as generally true might easily be misleading.[1]

It might be helpful to take up our analogy once again.[2] Suppose our imaginary manufacturer has built all but one of the cars required for the fleet when an assistant takes him on one side to point out that a machine very similar to the type they are now building was constructed by the firm some time ago as part of a cancelled order, and that it is now sitting uselessly in the showroom. Wouldn't that machine do as the final member of the fleet?

The manufacturer knows that it can be used in this way only if it meets the 'interchangeable parts' criterion, and he is faced with deciding whether the apparent differences between the left-over machine and the ones he has just built are superficial or substantial from this point of view. All I wish to remark is that in making his decision it will be his engineering knowledge that he will be drawing on. Knowing how cars work, it will be a relatively easy job for him to inspect the odd man out and determine whether its parts could replace and be replaced by those belonging to the fleet proper. Similarly, it is by his engineering knowledge—his knowledge of how the organism in question functions—that the biologist, I am suggesting, distinguishes between characteristics likely to be true of the species as a whole and those that may well be idiosyncratic.

This then accounts for the reliability of induction in certain circumstances within the species. But can species themselves be arranged into inductive groups—in other words, can higher *taxa* be natural? We now try to show that this is indeed a possibility. Once again the general thesis that natural higher *taxa* unite individuals (and thus the lower order groups they belong to) with similar genotypes holds true; and once again we must develop this hint and specify the nature and limits of the similarity in question.

---

[1] The radical differences between the reproductive and related organs of male and female within the same species presents a complication: it would be easy, on the suggestion we are proposing, to regard the rabbit's ovaries for example as playing a key role in the functioning of the animal's system as a whole, and thus conclude that it would be safe to think of all rabbits as having ovaries: but this kind of error would not be likely to survive a more complete understanding of how reproduction in the organism in question was achieved.

[2] I wish nothing of substance to hang on my analogies of course: they are meant as expository devices only.

There are two elements to my solution of this problem, and it will probably be clearest to separate them. The first represents an attempt to describe in the abstract one possible set of circumstances which would allow natural higher taxa to be set up; and as a second step I try to show that the mechanism of evolution has produced such a set of circumstances, so that in fact organisms can be so classified. One might say that I first set up a model, and then attempt to show that the model is based on axioms that reflect the facts so that we can accept it as a possible explanation of how natural higher taxa are possible.

We develop the model by considering a system of which the parts or components are themselves systems: *i.e.* a whole which functions in virtue of the ordered interaction of sets of components organised into sub-systems. If one thinks of a heating furnace as a system, the notion of a sub-system would be illustrated by the apparatus which supplies the fuel, and also by the mechanism by which the furnace is turned off and on as the situation demands. For the functioning of the system as a whole, the mechanism of each of its sub-systems is a matter of indifference. It is only necessary that each sub-system perform its allotted role, and if there are alternative ways in which that role may be performed it will not make any difference to the properties of the system as a whole which alternative is actually operative. In terms of our example, it doesn't matter whether the control is mediated mechanically of electrically, so long as each is as effective as the other in reflecting the heating demands of the environment for the temperature of which the furnace is responsible.

There would be a level of discussion, therefore, at which the properties and functioning of the system as a whole might be spoken of without any implications being made regarding the mechanisms by which the sub-systems performed their role. Moreover, it would be possible to have a group of systems which were identical as far as their overall organisation went, but which differed from each other in respect of the internal organisation of their sub-systems. The vital point for our present purpose is that such a group of systems would be 'inductive': if one knew that a certain individual system belonged to such a class as we are imagining, and one established some facts about this individual as a result of studying it, it would under certain circumstances be possible to infer that these facts held of the other members of the group. And the 'certain circumstances' would be these: when the facts established were facts about the properties of the system as a whole—at a level, that is to say, at which the internal organisation of the sub-systems was irrelevant.

It is easy to see, further, that what is true of systems with two levels of organisation—the system and the sub-systems that make it up—can be generalised to apply to multi-level systems, that is, to systems complicated by possessing one or more intermediate grades of organisation between the level of the system as a whole and the sub-systems of the case just discussed. For any given level one could imagine a group made up of systems identical from that level upwards, but otherwise showing variation. And such groups would be inductive: for in each case facts established in connection with a single member of the group, provided it concerned the right level of organisation, could be inferred to hold for the other members too. One can therefore imagine a hierarchy of inductive groups, new groupings being possible at each level of organisation. If we call the level of organisation at and above which members of the first order groups are identical the *first* level of organisation, then we shall say that second order groups will unite first order groups sharing a common organisation at the second level and above; third order groups will unite second order groups which have the same organisation at the third level and above; and so on. For each level of organisation there will be the possibility of inductive groupings, and as the level of organisation gets higher, so the groupings become more inclusive and so fewer. In brief, then: corresponding to the hierarchical organisation of the individuals concerned one might erect a natural hierarchical classification of them. So much for the model. How does it relate to what we know of organisms and their relationships?

Perhaps the picture of the organism as a system built out of a hierarchy of sub-systems according to the rigorously tidy pattern envisaged in the model is something of a simplification, yet I think the general approach is valid enough. An organisation *can* generally speaking be regarded as a hierarchy of systems, and we appeal to the work of J. Z. Young, for example, as vindicating this conception (see particularly the first chapter of his [1959]. And insofar as the relationship between the levels of organisation departs from the tidy hierarchical pattern, so far, on my own view, will the tidy hierarchical classification of the organisms in question run into anomalies.

But even if it is legitimate to regard organisms as hierarchies of systems we have not yet shown that they can be classified in the way indicated: for one can easily imagine a collection of individuals all with hierarchical organisation but all with quite *different* hierarchical organisations. At this point we must turn to the mechanism whereby the different species of living things have arisen: that is, the process of evolution. We shall see that it is this process which has resulted in the organisations of organisms bearing to each other the resemblances represented in the model. In other words, we are to show how evolution has generated a hierarchy of inductive groups of organisms which reflects the hierarchy of levels in the organisation of the individuals that compose them.

The central observation to make is that the huge variety of different species of organisms that we know of have evolved, over the millennia, from relatively very few, or even a single, progenitor-groups or group; and by a process, speciation, which sets very close limits on the size of the steps by which these ultimately extremely wide divergences can be achieved. This is because new species arise, typically, not as a result of a single individual developing a large variation and begetting a host of offspring but by partition of the parent species. (See *e.g.* Cain [1954], chapter 8.) A section of the latter typically becomes spatially isolated so that interbreeding with the remainder of the parent species fails to occur. The newly delimited gene pool then responds to selection pressures now peculiar to it and so the isolated population diverges from its parent group. Here then is the limitation on the degree of difference that can exist between parent and offspring gene pools: not only do the genotypes of the offspring species have to be mutually compatible in the way already specified (*above*, pp. 320–21) but at the point at which the spatial barrier becomes operative they must also be compatible with the genotypes of the parent population. For at that point parent and offspring populations are capable of interbreeding—only the spatial separation prevents it. In consequence fundamental changes in the structure and function of the evolving organism—that is, modifications to the higher levels of organisation—have to wait on the accumulation of modifications at less fundamental levels; as evolution proceeds, different modifications affecting the same level of organisation (which will occur in so far as the appropriate environments present themselves) generate a number of *taxa* of the same rank. Then, with the accumulation of a coherent set of modifications at one level, a modification of the level next above, and so a higher rank of *taxa*, becomes possible. Each higher category in a natural scheme of classification will thus correspond to a level of organisation in the hierarchy of systems of which the organisms to be classified consist, and the number of higher ranks it is legitimate to recognise will depend on the number of levels of organisation it is legitimate to recognise in the organisms concerned.

CONCLUSION

In outline our 'theory' of classification is then this. A natural classification will be one that classifies on the basis of genotypical similarity. The job we need a classification to help us with is the job of studying organisms scientifically. In particular we have pointed out that a basic division of individuals into inductive groups is an absolute precondition of investigating them scientifically. The secondary organisation of those basic groups into a hierarchy of wider groupings is not necessary in the same way; yet clearly it has the effect of making the scientific method more powerful in as much as it enables certain scientific findings (those concerning the appropriate level of organisation) to be applied beyond the confines of the basic group within which they were established. Why is it that classifying according to genotype should produce inductive groups? The answer is given by the elementary facts of genetics: a similar genotype, other things being equal, will produce a similar organism—and we took pains to define the similarity in question here. How do we set up groups of genotypical similarity? Of course there is no unproblematic method. But because the process of evolution has generated groups of the kind in question we shall move towards our goal by attempting to set up the appropriate kind of phylogenetic classification.

The substance of our discussion therefore amounts in a sense to a vindication of the phylogenetic approach characteristic of those who perhaps constitute today's mainstream taxonomic theorists. It is the rationale of such an approach that we have chiefly been attempting to clarify. An evolutionary classification is not self-justifying, as some writers seem to suppose: a taxonomist is trying to do more than produce a tool for the specialist in evolution. Those who go on to justify evolutionary groupings in terms of the fact that they constitute at the same time groups of genetic similarity present a much more satisfactory account, it seems to me. I simply take two philosophical paces beyond this and attempted to say, first, what precisely must be meant by 'similarity' in this context, and second, why it is that groups of genetic similarity in this sense are the sort of groups which are all-important for biological science.

University College, Cardiff

REFERENCES

BAMBROUGH, R. [1961]: 'Universals and Family Resemblances', *Proceedings of the Aristotelean Society,* **61**, pp. 207–22. (Page references are to reprint in G. Pitcher (*ed.*): *Wittgenstein, The Philosophical Investigations,* 1968, pp. 186–204.)

CAIN, A. J. [1954]: *Animal Species and Their Evolution.*

CAIN, A. J. [1959]: 'Deductive and Inductive Methods in Post-Linnean Taxonomy', *Proceedings of the Linnean Society of London,* **170**, pp. 185–217.

CAIN, A. J. [1962]: 'The Evolution of Taxonomic Principles', in G. C. Ainsworth and P. H. A. Sneath (*eds.*): *Microbial Classification,* pp. 1–13.

CAIN, A. J. and HARRISON, G. A. [1958]: 'An Analysis of Taxonomist's Judgement of Affinity', *Proceedings of the Zoological Society of London,* **131**, pp. 85–98.

CAIN, A. J. and HARRISON, G. A. [1958]: Phyletic Weighting', *Proceedings of the Zoological Society of London,* **135**, pp. 1–31.

CUVIER, G. L. C. F. [1817]: *Le Règne Animal.*

DUHEM, p. [1906]: *La Théorie physique, son objet et sa structure.* (Page reference is to the English translation of the second, 1914, edition: *The Aim and Structure of Physical Theory,* 1954.)

GILMOUR, J. S. L. [1951]: 'The Development of Taxonomic Theory since 1851', *Nature*, **168**, pp. 400–2.

GILMOUR, J. S. L. [1961]: 'The Mathematical Assessment of Taxonomic Similarity, Including The Use of Computers', *Taxon*, **10**, 97–101.

HEMPEL, C. G. [1965]: *Aspects of Scientific Explanation.*

HENNING, W. [1966]: *Phylogenetic Systematics.*

HESLOP-HARRISON, J. [1962]: 'Purposes and Procedures in the Taxonomic Treatment of Higher Organisms', in G. C. Ainsworth and P. H. A. Sneath (*eds.*): *Microbial Classification*, pp. 14–36.

HULL, D. L. [1967]: 'Certainty and Circularity in Evolutionary Taxonomy', *Evolution*, **21**, pp. 174–189.

HUXLEY, J. S. [1940]: *The New Systematics.*

MAYR, E. [1968]: *Principles of Systematic Zoology.*

MILL, J. S. [1843]: *A System of Logic.*

POPPER, K. R. [1935]: *Die Logik der Forschung.* (Page reference is to the expanded English edition: *The Logic of Scientific Discovery*, 1959.)

PRATT, V. [1972]: 'Numerical Taxonomy – A Critique', *Journal of Theoretical Biology*, **36**, pp. 581–92.

PRICE, H. H. [1953]: *Thinking and Experience.*

QUINE, W. V. O. [1969]: 'Natural Kinds', in N. Rescher (*ed.*): *Essays in Honour of Carl G. Hempel*, pp. 5–23.

SIMPSON, G. G. [1961]: *Principles of Animal Taxonomy.*

SOKAL, R. R. and SNEATH, P. H. A. [1963]: *Principles of Numerical Taxonomy.*

THOMPSON, W. R. [1952]: 'The Philosophical Foundations of Systematics', *Canadian Entomologist*, **84**, pp. 1–16.

WHEWELL, W. [1840]: *The Philosophy of the Inductive Sciences.*

YOUNG, J. Z. [1959]: *The Life of Mammals.*

Brit. J. Phil. Sci. **23** (1972), 329–346 *Printed in Great Britain*

Discussions

TWO POINTS IN THE THEORY OF STATISTICAL INFERENCE

This note discusses two points arising out of Hacking's review [1972] of Edwards's [1972]. The first concerns likelihood and arises direct from Hacking: the second concerns the fiducial argument, and arises from Edwards.

Hacking finds difficulty in the likelihood approach to statistical inference because with a single observation, say $x=2$, from a population in which x is normally distributed with unknown mean μ and unknown variance σ^2, we seem to obtain the estimates 2 for μ and 0 for σ^2. Hacking accepts 2, with some hesitation, as the most plausible value for μ but he finds the suggested value 0 for σ^2 'monstrous'.

I have argued elsewhere [1967] that the singularity in the likelihood function

$$L=(1/\sigma)\, e^{-\frac{1}{2}}((\mu-2)/\sigma)^2$$

at the point $(\mu,\sigma)=(2,0)$ is removed if we bear in mind that x will in practice have been observed only to a finite degree of accuracy—as, say, 2.00, rather than as 2.000 . . . *ad infinitum*. This point is also made by Edwards, and so presumably Hacking does not feel that this disposes of his difficulty. I therefore present a complementary argument in which, merely for the sake of simplicity, we ignore the finiteness of the accuracy of x.

The fact is often overlooked that, except when $\mu=2$, the function L, considered as a function of σ, is what mathematicians call 'quasi-analytic' at the origin. This means that at the point $\sigma=0$, the function L *seems* well-behaved, in that it possesses a Taylor series, but function is not equal to its Taylor series. The technicalities need not concern us now, except to warn us that there may be pitfalls in standard methods of finding maxima, such as setting the derivatives of L equal to zero.

If we put μ equal to 2, the exponent in L vanishes and we simply have

$$L=1/\sigma$$

which grows infinitely large as σ approaches 0, suggesting the 'monstrous' estimate $\sigma=0$. But if, for instance, we set μ a little different from 2, say 2.1, then we have

$$L=(1/\sigma)\, e^{-(0.005/\sigma^2)}$$

and this has its ***minimum*** value at $\sigma=0$. and its maximum at $\sigma=0.1$. In other words, we get the 'monstrous' value 0 for σ only after taking the value 2 for μ.

The likelihood function is therefore telling us that separate estimate of μ and of σ is impossible. We may speak of the ***pair*** (2.0) as the most plausible ***pair*** of values for (μ, σ), but we cannot discuss the most plausible value for σ by itself.

This dependence of the most plausible value for σ on the value adopted for μ is always present, even with larger samples of $n > 1$. If the usual unbiased estimates are $\bar{x}$ and s^2 the most plausible value for σ^2 is

$$((n-1)/n)s^2+(\bar{x}-\mu)^2.$$

When n is large, x is unlikely to be far from μ, so that the second term here becomes small and the most plausible value for σ^2 is unlikely to be far from s^2. Thus we have developed the habit of thinking of s^2 as giving an estimate of σ^2 by itself; but strictly speaking this is only an approximation. This approximation goes badly wrong when $n=1$, since then it is the first term which vanishes and the term $(\bar{x}-\mu)^2$ is the only part left.

But can we sensibly speak of the pair (2, 0) as the most plausible value for the pair (μ, σ)? I think we can. Let us imagine that we have a measuring instrument which we know to be unbiased, but about whose accuracy we know nothing—for all we know the instrument may be completely accurate. We can read the instrument to as many decimal places as we please. We have a force which we know to vary as the inverse αth power, where α is a positive real number; there is a theory which asserts that α is exactly 2. We measure α with our instrument and we obtain the reading 2.000000000 . . . to a million decimal places. Such a result would lead me to suspect strongly that the instrument was highly accurate and that the theory was correct. Of course, in practice, one would try to obtain at least one more reading; but to keep to the cases considered we have to suppose that this is impossible.

Now for Edwards's point about the fiducial argument. It would, of course, take too long to discuss the merits of the fiducial argument here. I am now concerned only with Edwards's objection to it, not with the other objections that have been raised from time to time.

Edwards considers a situation where we have three quantities X, Y, and E, each known to take only the values ± 1. It is also known that $XY=E$, and that the probability is $\frac{1}{2}$ that $E=+1$. If, now, we are given the further information that $Y=+1$, then it follows that the probability is $\frac{1}{2}$ that $X=+1$; while if, instead, we are given the further information that $Y=-1$, the probability is still $\frac{1}{2}$ that $X=+1$.

Edwards argues that since the distribution of X is the same whichever value Y takes, an observation on X must be completely uninformative about Y. So it seems paradoxical that, from the *a priori* information specified above, together with the statement that this is all that is known *a priori*, the fiducial argument would seem to allow us to infer, from an observation $X=+1$, that the probability is $\frac{1}{2}$ that $Y=+1$. Either, says Edwards, we have here a justification of the principle of indifference (which would allow us to say, in the absence of further knowledge, that $Y=+1$ and $Y=-1$ are equally probable), or the fiducial argument is fallacious. How can knowing the result of a completely uninformative experiment convert the unknown (or non-existent) probability that $Y=+1$ into a known probability of $\frac{1}{2}$?

It seems to me that an answer lies in the vagueness of the phrase 'completely uninformative'. Let us define 'absolutely uninformative'. by saying that proposition A is absolutely uninformative about a parameter θ if and only if *for all possible B*, knowledge that A is true added to knowledge that B is true leaves the information about θ in B unchanged. Symbolically, if we denote by *Info* $(\theta{:}K)$ the information we have about θ when the knowledge we have is that K is true, we say that A is absolutely uninformative about θ if and only if,

$$\textit{for all } B, \ \text{Info}\,(\theta{:}\ A \text{ and } B) = \text{Info}(\theta{:}B).$$

Now A and B can each be completely uninformative, in Edwards's sense, about

θ, while A and B together may completely determine θ. This is proved by the following simple example: We are given that the only possible values for the triple (x, y, θ) are $(0, 0, 0)$, $(1, 1, 0)$, $(0, 1, 1)$ and $(1, 0, 1)$, and each of these has probability $\frac{1}{4}$. A is the proposition '$x=+1$', while B is the proposition '$y=+1$'. Without knowing either A or B, the values 0 and 1 are equally probable for θ. The same remains true if we know A to be true, and also if we know B to be true. Thus A and B are each completely uninformative about θ, in Edwards's sense. But if we know *both* A and B to be true, then we know that θ must be 0; thus neither A nor B is absolutely uninformative about θ.

Applying these ideas to Edwards's example, we see that '$X=+1$' is completely but not absolutely, uninformative about Y since knowledge that $X=+1$ added to knowledge that $E=+1$ completely determines Y as $+1$. To obtain a general justification of the principle of indifference from applying the fiducial argument here, we would need to have X absolutely uninformative about Y, which is not the case. While it seems paradoxical that a completely uninformative experiment should enable us to say something about Y that we could not say before, it seems equally paradoxical that putting two completely uninformative statements together should give us complete information; and yet the example quoted shows that this latter must be so. The conclusion must be, that the phrase 'completely uninformative' must not be taken too literally.

G. A. BARNARD
University of Essex

REFERENCES

EDWARDS, A. F [1972]: *Likelihood. An Account of the Statistical Concept of Likelihood and its Application to Scientific Inference.*

HACKING, I. [1972]: 'Likelihood', *British Journal for the Philosophy of Science*, **23**, pp. 132–7.

PROPENSITY THEORIES OF PROBABILITY UNSCATHED: A REPLY TO WHITE

1 Alan White's recent article, ([1972]), contains a number of errors of exegesis of a note by C. S. Peirce ([1910]), which should not be left uncorrected, as well as a number of other errors which render his criticisms of attempting to analyse probability in terms of liability beside the point. First the exegetical errors.

2 In calling Popper's propensity theory of probability (see his [1957], [1959] and [1967]) a 'resuscitation' of Peirce's (p. 35), White failed to take proper account of two significant differences between the theories. Peirce ascribes his 'would-be' as a property of objects regardless of environment—in this he is followed by Mellor (see his [1969], [1971])—while Popper explicitly makes them relative to environment. This is not simply a matter of the conventions of language use, as Mellor suggests, but rather, and more importantly, a matter of what we want to assert, as scientists, about the world. Do we wish to say that the disposition of

falling to the ground is a disposition of an object, as such, regardless of its environment? Would we not rather say that it was a disposition relative to a gravitational field? Secondly, Popper's theory, but not Peirce's, is aimed to make sense of single-case probabilities.

White's second exegetical mistake is to ascribe to Peirce the intention to *analyse* the notion of probability (p. 35). As far as I can tell, Peirce's [1910] attempts no such analysis. Certainly Peirce does not describe what he is doing as analysis. Rather he says he is aiming to *define* probability ([1910], pp. 404, 405, 409) a word 'that we, all of us, use . . . with a degree of laxity which corrupts and rots our reasoning' (p. 405). Peirce's method is first to distinguish probability from likelihood and from plausibility by characterising the latter terms (pp. 406–8); and then to define the meanings of probability statements by giving synonyms, one of which he coined for the purpose. This is not analysis. Indeed, it is questionable whether we are called upon to analyse probability, construed as a propensity, any more than we are called upon to analyse gravitation, construed as a field of force. Some positivists attempt analyses of probability in terms of the relative frequencies of different kinds of outcomes of trials on a chance setup, but their analyses, one suspects, are aimed to eliminate probability as an unanalysed primitive, in favour of discrete events or inert properties of particular objects (being head uppermost, say) as primitive. What Peirce's theory aimed to do, by contrast, was to suggest probability had roots in the qualities of the relevant physical objects.

Thirdly, White misrepresents what Peirce calls a 'habit'. White claims that a habit differs from a liability in the same way, 'what does happen when' differs from 'what would happen if'. (p. 35), which is an infelicitous contrast in view of Peirce's identification of habit with 'would-be'. Indeed Peirce is very careful not to confuse habit in his sense with habit in White's sense, that is, repetitious behaviour. (Peirce [1910], pp. 409–10, 412).

3 Perhaps a more serious shortcoming is White's attempt to stifle new developments in physical theory by requiring all words to retain what he thinks is current English usage. He writes 'Unfortunately for the (propensity) theory . . . probability is not *in fact* attributed to such things as coins, dice . . . *etc.* It is not the probability of the coin, but of the coin's falling heads, that is in question' (p. 38, my italics). Thus, with one sweep of his bold 'in fact', White slays all propensity theories of probability! One must protest!

In the first place, one may challenge White's account of common usage. As Popper remarked, 'the word "probability" can be used perfectly properly and legitimately in dozens of senses, many of which, incidentally, are incompatible with the formal calculus of probability' ([1967], p. 31). (See also Popper [1935], Carnap [1945]; and Bunge [1967*a*], [1967*b*].) White appears to be unable to shake off frequentist predilections sufficiently to think probability could carry any other sense than that of the chance of certain outcomes of trial runs on a chance set-up. Of course, I am not suggesting that 'probability' does not carry the sense of the chance of certain outcomes of trial runs on a chance set-up. On the contrary, I insist that it may carry that sense. The frequency interpretation of the probability calculus (for example, see Von Mises [1919*a*], [1919*b*]) allows us to ascribe probability as a property of the coin's falling heads relative to other possible outcomes. Nor am I suggesting that propensity theory is incompatible with that interpretation of the probability calculus. Far from it. (Compare Bunge [1967*a*], p. 91, where Bunge proposes a physical probability theory which is the union of three interpretative models). My point is that under an interpretation other than the frequency interpretation (say a propensity interpretation) probability may be ascribed otherwise than as a property of a coin's falling head uppermost (say as a dispositional property of the coin in a gravitational field *etc.*) This leads me to my second point that common usage (even if White is right about what is common usage) should not be an arbiter over propriety of scientific usage, since at least one aim of science is to reform knowledge, and hence to reform common sense and common usage. This point is relevant to the debate over propensities since from one point of view, propensity theory is a clear, consistent interpretation of a mathematical formalism (the calculus of probabilities) introduced to solve a problem in theoretical physics, especially quantum theory. (See Popper [1957].) Incongruities between common usage and subsequent usage in physics are incidental rather than crucial to the success of propensity theory.

In the third place, even though the chance of falling head uppermost may be a property of falling head uppermost, the *propensity* to fall head uppermost is certainly not a property of falling head uppermost. Both Peirce and Popper, who introduced propensity theories independently, intended to introduce a novel interpretation of probability. Both argued that the probability of a die falling with a particular face uppermost was, for some purposes, better construed as *a property of the die*, (Peirce) or *of the die relative to an environment* (Popper). In this matter, all the propensity theorists I refer to agree, whatever else they disagree about.

In sum, White fails to distinguish between probability construed as a frequency and probability construed as a propensity. The distinction between these construals is of some importance, especially regarding the question of assigning probabilities to single cases, as Popper argues (see his [1967]; compare Giere [1973]), even though the connection between them is also of importance where the propensity in question is a disposition to produce outcomes of a certain kind with a certain frequency.

4 White's most serious mistake, serious because it ruins his argument, is to identify propensity with liability. To be sure, Popper is to blame for introducing the phrase 'is liable to' as a partial explication of propensity (cf. his [1957], p. 67). But White is to blame for taking the synonymity of liability and propensity seriously, because liability is referred to nowhere else in Popper's writings on propensity theory. Usually Popper uses 'tendency' or 'disposition' as synonyms. (See, for example, Popper [1957], pp. 67, 68, 70; [1959], pp. 30, 31, 35, 36, 37; [1967], p. 32.)

White is correct to notice the categorial difference between liability and likelihood, and quite correct to affirm 'one cannot analyse the notion of being likely in terms of the notion of *being liable*' (p. 43). But the declared purpose of his paper was to show the propensity theory of probability to be mistaken (p. 35). Remarks about liability are not at all to that purpose. White's criticisms concerning liability do not hold for tendency or disposition, save that likelihood is not to be analysed in terms of tendency or disposition, either. But then propensity proponents do not use these concepts to analyse 'probability' but to interpret it. Propensity theory of probability remains unscathed from White's attack, since precious little remains of his argument if propensity theory does not imply either that probability in the sense of chance is to be analysed in terms of propensity or that propensity is to be understood as liability.

5 One final remark. It should not be surprising if proponents of propensity theory of probability differ. There is not simply one theory, as White's title suggests; there are at least eight significantly different theories, varying in their acceptability, ascribable severally to Peirce ([1910]), Braithwaite ([1953]), Popper ([1957]), Hacking ([1965]), Bunge([1968], [1969]), Mellor ([1969], [1970]), Giere ([1973]) and myself ([1972], [1973]), aside from a 'qualified propensity theory' such as Levi's ([1967]) and a caricature of a propensity theory such as Sklar ([1970]) criticised, but which does not belong recognisably to any of the authors mentioned. Incidentally, in none of these theories (save Sklar's caricature) is chance analysed in terms of propensity and in none (save Popper's slip of the pen) is propensity identified with liability.

TOM SETTLE
University of Guelph

REFERENCES

BRAITHWAITE, R. B. [1953]: *Scientific Explanation.*

BUNGE, M. [1967*a*]: *Foundations of Physics.*

BUNGE, M. [1967*b*]: *Scientific Research I.*

BUNGE, M. [1968]: 'Philosophy and Physics' in R. Klibansky (*ed.*): *Contemporary Philosophy*, volume 2, pp. 167–199.

BUNGE, M. [1969]: 'What are physical theories about?', in N. Rescher (*ed.*): *Studies in the Philosophy of Science* (*American Philosophical Quarterly Monograph, number 3*) pp. 61–99.

CARNAP, R. [1945]: 'Two Concepts of Probability', *Philosophy and Phenomenological Research*, 5, pp. 513–532. Reprinted in H. Feigl and W. Sellars (*eds.*) *Readings in Philosophical Analysis*, 1969, pp. 330–348.

GIERE, R. N. [1973]: 'Objective Single Case Propensities and the Foundations of Statistics', forthcoming in P. Suppes (*ed.*) *Proceedings of the Fourth International Congress on Logic, Methodology and Philosophy of Science.*

HACKING, I. [1965]: *Logic of Statistical Inference.*

LEVI, I. [1967]: *Gambling with Truth.*

MELLOR, D. H. [1969]: 'Chance', *Aristotelian Society Proceedings, Supplementary Volume*, 43, pp. 11–36.

MELLOR, D. H. [1971]: *The Matter of Chance.*

PEIRCE, C. S. [1910]: 'Notes on the Doctrine of Chances', *Popular Science Monthly*, 44. Reprinted in C. Hartshorne and P. Weiss (*eds.*): *Collected Papers of Charles Sanders Peirce, Volume 2*, 1932, pp. 404–414.

POPPER, K. R. [1935]: *Die Logik der Forschung.*

POPPER, K. R. [1957]: 'The Propensity Interpretation of the Calculus of Probability and the Quantum Theory', in S. Körner (*ed.*) *Observation and Interpretation in Philosophy and Physics*, pp. 65–70.

POPPER, K. R. [1959]: 'The Propensity Interpretation of Probability', *British Journal for the Philosophy of Science*, 10, pp. 125–42.

POPPER, K. R. [1967]: 'Quantum Mechanics without "The Observer" ', in M. Bunge (*ed.*): *Quantum Theory and Reality*, pp. 7–44.

SETTLE, T. W. [1972]: 'Induction and Probability Unfused', forthcoming in P. A. Schilpp (*ed.*): *The Philosophy of Karl R. Popper.*

SETTLE, T. W. [1973]: 'Are some propensities probabilities?', forthcoming in P. Suppes (*ed.*): *Proceedings of the Fourth International Congress on Logic, Methodology and Philosophy of Science.*

Sklar, L. [1970]: 'Is probability a dispositional property?', *Journal of Philosophy*, 67, pp. 355–366.

Von Mises, R. [1919*a*]: 'Fundamentalsätze der Wahrscheinlichkeitsrechnung', *Mathematische Zeitschrift*, 4, pp. 1–97.

Von Mises, R. [1919*b*]: 'Grundlagen der Wahrscheinlichkeitsrechnung', *Mathematische Zeitschrift*, 5, pp. 52–99.

White, A. R. [1972]: 'The Propensity Theory of Probability', *British Journal for the Philosophy of Science*, 23, pp. 35–43.

A COMMENT ON GILLIES'S FALSIFYING RULE FOR PROBABILITY STATEMENTS

There is an obvious objection to the approach of Dr Gillies, *viz.* to divide the family of probability measures on the real line into two classes, those which are liable to falsification and those which are not.

The objection is due to the fact that any statistic is essentially unique at most up to monotone transformations. Hence if we get a statistic Y which is t-distributed we will do equally well with $Z=F(Y)$ where $F(y)=prob(Y\leqq y)$, *i.e.* F is the distribution function of Y. Now Z is uniformly distributed over the interval (0,1) and not liable to falsification although Y is.

Conversely if Z is uniformly distributed over (0,1) we can transform Z to any continuous distribution by using the inverse of the distribution function.

Moreover a similar approach is often used by 'the practical man'. If he has got the observation y of a statistic Y, he will calculate the probability p to get 'something worse' than y, *i.e.* in the case of a two-sided t-test $p=P(|Y|>|y|)$. He will then reject the hypothesis if $p\leqslant k_0$, the level of significance. It is well-known that p in quite general cases is an observation of a random variable P with uniform distribution over (0,1). So even in practice you will often transform the problem into a 'non-falsifiable distribution'.

JACQUES DE MARÉ
Department of Mathematics
Lund Institute of Technology, Sweden

REFERENCE

Gillies, D. [1971]: 'A Falsifying Rule for Probability Statements', *British Journal for the Philosophy of Science*, 22, pp. 231–61.

REPLY TO DE MARÉ

J. de Maré does not state my position correctly. He writes: 'There is an obvious objection to the approach of Dr Gillies to divide the family of probability measures on the real line into two classes, those which are liable to falsification and those which are not'. However I do not in fact try to divide the probability measures on the real line into two classes. What I do do is try to divide the distributions of particular random variables ξ into those which are falsifiable and those which are not. I even make the further restriction that the distributions in question should be either discrete or continuous.

To see that those two things are not the same, let us suppose we have a probability measure on the real line R. In general it will be possible to define a large number of random variables ξ, η, . . . with R as domain, some of which will have falsifiable distributions and some of which will not. Thus we cannot, as it were, carry back the falsifiable/non-falsifiable distinction to the underlying measure. Indeed we can only do so if we suppose that the measure is given canonically by some particular random variable, and there is no need in general to suppose this.

My position would, I believe, have been made clearer had I spoken of dividing *distribution functions* into two classes. J. de Maré's objections in no way prevent us dividing such functions into those which are falsifiable and those which are not, along the lines suggested in my paper—provided that the distributions in question are either continuous or discrete.

D. A. GILLIES
Chelsea College of
Science and Technology
University of London

FREE WILL IN A MECHANISTIC UNIVERSE? AN EXTENSION

Introduction.
1 *The Definitions.*
2 *Analysis of Definitions.*
3 *Implications of the Validity of P.*
4 *Discussion of the Definitions.*
Conclusions.

Introduction.

In a recent paper (Landsberg and Evans [1971], to be referred to as *LE*) we introduced three definitions of free will and investigated to what extent they may be expected to exist in a mechanistic universe. We concluded that our weaker forms of free will, denoted by *D2* and *D3* (see *below*), may be present in certain cases but that the stronger form *D1* (which we gave reasons for preferring) is never to be expected. We utilised arguments based on predicted brain states of an agent A, the predictions being communicated to A in advance. We did not discuss cases where the prediction is kept secret from A, an omission which will be made good in the present paper.

In this paper we introduce two new definitions of free will and investigate all five definitions by a new uniform method. In this way we cover the cases (*a*) and (*b*) of *LE* (as set out in the summary on p. 353 of that paper.) While we do not cover again the cases (*c*) and (*d*) of the earlier paper, we investigate the question of predictions which are not communicated to A, by means of the new definitions *D5*. Our conclusion is that again only the weaker forms of free will occur, and that the introduction of secret predictions does not improve the prospect of free will in a mechanistic universe.

It was perhaps a merit of the precision which we sought to give to the definitions of free will in *LE*, that MacKay [1971] has now presented a more explicit necessary condition for the existence of free will. This is welcomed, since he has pursued different variants of related arguments in his publications (MacKay [1965] [1967], and see there for other references), and misunderstandings seem to have occurred. In order to avoid possible misunderstanding we shall therefore, be careful to argue entirely in our own terms, except to note here MacKay's thesis.

M: There are at least some of *A*'s future brain states for which it is impossible to formulate (even secretly) a complete prediction having a demonstrable claim to *A*'s unconditional assent.

One may ask what unconditional assent means operationally. If one chooses the interpretation 'that *A* would be correct to agree with the prediction and wrong to disagree with it' then *M* asserts the presence in some situations of (*D4*) and (*D5*) free will. The arguments of section 5, *below*, for regarding the definitions *D4* and *D5* as too weak to be useful are, therefore, on this interpretation, also arguments for denying that *M* implies free will in any useful sense.

It is only to this extent that this paper is related to MacKay's note ([1971]). The points of detail raised there will not be taken up, but the reader is referred to Good [1971] and Watkins [1971].

1 *The Definitions.*[1]

In *LE* we considered a situation in which a scientist *S* had a correct and detailed description of the brain state of an agent *A*, at time t_0. The controlled conditions under which *A* is held at least up to time t_2 enabled *S* to predict *A*'s brain state at time t_2. The possibility of communicating this prediction *P* to *A* in an appropriate language *L* at an intermediate time t_1 was also envisaged. This led us to introduce three definitions of free will. Emphasising the differences between these definitions by italics, and inserting the times explicitly (they were clear from Fig. 1 in *LE*), these definitions were

D1 It is impossible *for any observer* (other than *A*) to possess a complete and certain prediction of one of *A*'s future brain states.

D2 It is impossible *to provide A* at time t_1 with a complete and certain prediction of *A*'s brain state at the later time t_2.

D3 It is impossible *to provide A* at time t_1 with a complete and certain prediction of *A*'s brain state at the later time t_2 *with which A would be correct to agree at* t_1.

In *D2* and *D3* it is understood by us, though it is not essential to the argument, that all rational observers would believe a complete and certain prediction; however, only *A* is mentioned in these definitions since it is only for *A* that logical difficulties can arise. We also showed in *LE* that these definitions are in order of increasing weakness, *i.e.* as one passes down the list additional situations conforming to the definition become available.

Although we gave reasons for regarding *D2* and *D3* as too weak for what we would regard as valid definitions, it is of interest to continue the list to include the even weaker definitions *D4* and *D5*, as follows:

---

[1] We have chosen to express our Definitions in the form of conditions (*Dr*) so that (*Dr*) is necessary and sufficient for the existence of (*Dr*)-free will in a particular situation.

(*D4*) It is impossible *to provide A* at time t_1 with a complete and certain prediction of *A*'s brain state at the later time t_2, *such that A would be correct to agree with the prediction at* t_1 **and wrong to disagree with it at t_1.**

(*D5*) It is impossible *for any observer* to possess at time t_1 a complete and certain prediction of *A*'s brain state at the later time t_2, *which is not intended to be communicated to A at time* t_1, *but nevertheless would remain valid if A were to hear of it at* t_1, *so that A would be correct to agree with the prediction at* t_1 **and wrong to disagree with it at t_1.**

It will be seen that the bold print in *D4* is essential for what follows, since the additional restriction which it represents makes the impossibility envisaged in the definition easier to be realised in a given situation. The bold print in *D5*, on the other hand, is not essential. It is in the nature of an interpretation of the preceding phrase, and is designed to bring out certain formal similarities with *D3* and *D4*.

This extended spectrum of definitions can be analysed with reference to certain characteristics of the prediction *P*. These are

C1: *P* is valid if *A* agrees with it at t_1.

C2: *P* is valid if *A* disagrees with it at t_1.

C3: *P* is valid if *A* does not hear of it at t_1.

We allow the possibility of communicating the prediction *P* to the agent *A* only at time t_1, in order to attain clarity in the exposition. Since each of *Cn* may be true or false in a given situation, there are eight cases to be considered. In

TABLE 1

Free will implications of eight classes of predictions

(T denotes truth and F denotes falsehood)

Number of case	*If* a prediction is possible with property			*Then* free will exists according to definition					Corresponding cases of *LE*
	C1	*C2*	*C3*	*D1*	*D2*	*D3*	*D4*	*D5*	
1	T	T	T	F	F	F	F	F	*A* agrees, disagrees or does not hear of *P*
2	T	T	F	F	F	F	F	T	*A* hears of *P*. He agrees or disagrees
3	T	F	T	F	F	F	T	T	*A* does not hear of *P* or agrees with *P*
4	T	F	F	F	F	F	T	T	*A* agrees with *P*. Case (*a*) of *LE*
5	F	T	T	F	F	T	T	T	*A* does not hear of *P* or agrees with *P*
6	F	T	F	F	F	T	T	T	*A* disagrees with *P*. Case (*b*) of *LE*
7	F	F	T	F	T	T	T	T	*A* does not hear of *P*
8	F	F	F	T	T	T	T	T	No prediction is possible. Case (*d*) of *LE*.

each case one must determine which of the five types of free will exists. The results are given in Table 1, where we have adopted the convention that as one goes down the list more free will definitions are satisfied. One may say that, since it is satisfied only in case 8, $D1$ is the strongest definition of free will. Since $D5$ is most easily satisfied, it represents the weakest of our definitions of free will.

The last column of Table 1 is discussed in section 4.

The meaning of an 'F' entry in the table needs some discussion. Formally, an 'F' entry for Cr (r=1. 2 or 3) means that a prediction is *not* possible with property Cr. However, we assume (except in case 8) that a prediction P *is* possible, so that an 'F' entry for Cr means only that P does not have the property Cr, *i.e.* that the *proposition* Cr is false for P. The phrase 'Cr is false' will be used in the above sense. Even in case 8 one can suppose a prediction P to exist which lacks all three properties, but such a prediction would be simply invalid and so of no interest.

Suppose that for a particular prediction P, $C1$ is false (cases 5 to 8). Then P does not have the property $C1$, so that

A's agreement implies the invalidity of P. (3.1)

By the well-known rule that (V implies U) is equivalent to (not U implies not V), (3.1) is equivalent to

the validity of P implies that A does not agree with P

or more explicitly

the validity of P implies that A either disagrees with P or does not hear of P. (3.2)

Similar inferences can be drawn from the falsity of $C2$. Finally, if $C3$ is false, one can infer

P is invalid if A does not hear of P (3.3)

or equivalently

The validity of P implies that A hears of P. (3.4)

2 *Analysis of Definitions.*

For case 1 P is valid whether of not A hears of it and, if he hears of it, whether or not he agrees with it. Such a situation is plausible only if A is under some unusual constraint (brain damage, hypnosis, *etc.*) and it is reasonable that free will is then absent be all the definitions. In case 8 no valid prediction exists and accordingly free will is present by all the definitions.

The differences between the definitions are brought into play in the intermediate cases. In case 2, P becomes invalid if A does not hear of it, and the impossibility of a prediction which is valid whether or not A hears of it leads to the existence of $D5$-free will. In case 3, P becomes invalid if A hears of P and disagrees with it. In other words, A would be ***right to disagree*** with P. Under these conditions it is in fact impossible to provide A with a prediction as specified in $D4$, and one has $D4$-free will. However, ***A would also be correct to agree*** with P, and the impossibility stipulated in $D3$ is therefore absent, and $D3$-free will does not exist. It is therefore in case 3 that the phrase in bold print in $D4$ is first used. In case 4, P is valid only if A agrees with it and is invalidated if A disagrees or does not hear of the prediction. Hence the free will situation is as for case 3 In case 5, a valid prediction exists provided A does not hear of it, or if he hears of it, he disagrees with it. If he agrees with P he invalidates it. Thus $D3$-free will is now added to $D4$- and $D5$-free wills. In case 6, a valid prediction P exists provided A disagrees with it, and $D3$- to $D5$-free wills hold. In case 7, P is valid only if A does not hear of it, and this leads to $D2$-free will.

It will be observed that the free will implications of cases 3 and 4 are the same, even though in case 3 $C3$ is valid, and in case 4 it is invalid. This is due to the fact that the possibility of secret predictions P (which are the subject of $C3$) does not violate $D2$, $D3$ and $D4$ free wills. In particular, $D3$- and $D4$-free will, whose validity is at issue in these cases, are not affected by the validity of $C3$. The free will implications of cases 5 and 6 are also the same, and the reason for this is similar: $D2$-and $D3$ free wills are at issue in these cases and are unaffected by the validity of $C3$.

3 *Implications of the Validity of P.*

To the purely logical possibilities discussed in section 3 we now add the important consideration that the prediction P had been validly deduced from a valid scientific theory. Any inference in the Table which imply that P is invalid are therefore inadmissible. If one applies this to cases when $C1$ is false, one finds that (3.2), not (3.1), is the only admissible interpretation of '$C1$ is false'; similarly (3.4), not (3.3), is the only admissible interpretation of 'C3 is false'.

Bearing this additional restriction in mind, one may ask in each of the eight cases: Is it consistent that A agrees with P, or that A does not hear of P, *etc*? The last column of Table 1 shows the only possibilities which may be consistently entertained in each of the eight cases. In case 8 no result is consistent with the validity of the prediction P. For if A agrees with P, and P is valid, then $C1$ is true, which is not allowed. If A disagrees with P, and P is valid, then $C2$ is true, which is also not allowed. Lastly, if A does not hear the prediction, and P is valid, then $C3$ is true, which is not allowed.

Several results of *LE* arise from the present study as special cases, as will now be shown. In *LE* we considered only cases in which the prediction is communicated to A. However, we also envisaged the possibility that several consistent predictions can result from theory a in a given situation. We do not consider this case (*c*) separately here. The other cases considered in *LE* are obtained by requiring $C3$ to be false. Then case 4 covers (*ai*) and (*aii*) of *LE* and its only consistent interpretation is (*ai*), namely, that A hears of P, as discussed in *LE*, p. 351. Case (*aii*) is physically inadmissible because it implies an error in the prediction. Case 6 covers (*bi*) and (*bii*) of *LE*. Its interpretation is, in accordance with (*bi*), that A disagrees with P. A's belief of P would contradict P and is physically inadmissible, as discussed in *LE*, p. 351. Case 8 corresponds to case (*d*) of *LE*, when no self consistent prediction is possible. Case 2 was not discussed in *LE*, and corresponds to a prediction which remains valid whether or not A believes it.

4 *Discussion of the Definitions.*

We first criticise the definition $D5$, by considering two parallel situations S^1, S^2, in each of which A has been put under a rigid post-hypnotic suggestion, of which he is unaware, by the scientist S. In S_1 the suggestion is that, whatever is said to him at time t_1, he will perform an action X at t_2. In S_2 the suggestion is that he will perform X at t_2 if, and only if, the prediction P: 'you will do X at t_2' is communicated to him at t_1. We suppose further that for technical reasons P is the only prediction available to the experimenter at t_1.

It can now be seen that situation S_1 is an example of case 1 of the Table, allowing A no free will by any of the definitions while S_2 is an example of case 2 and allows (*D5* free will (because that truth of P is in this case not independent of A's hearing of it). In both situations, however, A's action at t_2 has been chosen for him by S, and is causally independent of his conscious intentions in the period (t_1, t_2). We consider that no definition which ascribes free will to A in these circumstances can be accepted.

The definition *D4* will now be criticised. As was pointed out in *LE*, p. 351, the calculation which leads to prediction P of A's brain state at t_2 proceeds by following the development of A's brain state in time from t_0 to t_2, and therefore contains implicitly a prediction P' as to A's brain state (and, in particular whether or not he will agree with P) at t_1. P and P' are both valid predictions. For instance, if P corresponds to case 4 of the table, P' must include the prediction that A will agree with P at t_1. A possesses *D4* free will in this case because, *if* A were to disagree with P at t_1, the calculation leading to P would be false as regards A's brain state at t_1, and presumably therefore P itself would be false; A would be right (not wrong) to disagree with P. The inference can, however, be reversed; since P and P' are both valid predictions of a valid theory, A's brain state at t_1 is correctly described by P' and he *will not* disagree with P. A possesses *D4* free will only in virtue of a logically possible event (his disagreement with P at t_1) which is physically impossible in a mechanistic universe under the given experimental conditions.

The definition (*D3*) is subject to the same criticism as (*D4*). Cases (5) and (6) of the Table arise for a valid prediction only if A's *disagreement* with P at t_1 is implicit in the prediction, and therefore physically determined in advance. A possesses *D3* free will in these cases only because P would be invalidated if a *Physically impossible* event (A's *agreement* with P at t_1) were to occur.

A different criticism of (*D3*), (*D4*) and (*D5*) can also be made. Consider an agent A who is predisposed for extra-logical reasons (paranoia, hatred of the experimenter, *etc*) *always* to disagree with whatever prediction P is communicated to him. In ordinary language, A is *not free to agree with P*, yet, in cases (5) and (6), he possesses *D3*, *D4* and *D5* free will. It appears than that all three definitions are too weak for the everyday meaning of the term 'free will'.

If these arguments are accepted, it remains only to consider (*D2*) and (*D1*). The critical case here is case (7) of the table, where a prediction P exists which is valid if A does not hear of it in advance. In this situation A is safe from the disconcerting experience of being presented with a correct prediction of his future brain state, and therefore possesses *D2*-free will. On the other hand we may assume that A knows that the prediction P exists and can, in principle, be correctly believed by the scientist S and any number of detached observers.

A may now say 'I understand that a correct prediction P of my future brain state exists. Therefore I have no free will'. He would then be adopting *D1*. He may say, alternatively 'P is not known to me and since it has been worked out on the assumption that I do not hear of it, I can choose my attitude to it as I wish and I have *D2*-free will in this respect.' Since all these reactions are in principle predictable by the scientist, this free will is of a subjective type: only the agent A himself can assume that he has it.

We adhere to the view that *D1*-free will is the only reasonable objective definition, and this is absent in all the cases investigated. But in the case of secret predictions belonging to case 7 an agent may reasonably adopt either *D1* or *D2* as free will definitions. In the latter case he will consider himself to have this type of free will.

Conclusions.

The analysis of five types of free will has in this paper been shown to lead to eight logically distinct cases, seven of which can arise when a specific prediction P about the brain state of an agent A exists. Although we prefer the strongest definition *D1* the next strongest definition *D2* is not unreasonable and leads to a subjective type of free will if the prediction belongs to case 7.

With each prediction considered here one can associate the number of the case in the Table to which it belongs, and hence various types of free will. We now consider the situation which arises if several predictions which belong to different cases of Table 1 apply to one and the same agent.

For example, if an agent A is subject to a prediction P belonging to case 2, and also to a prediction P' belonging to case 3, then the existence of P alone is sufficient to establish that A has at most *D5* free will. The existence of a second prediction P', which alone would allow *D4* and *D5* free will, does not alter the situation; A has *D5* free will only. Thus the free will situation is determined by the prediction which allows the *least* freedom.

Since the free will situation is determined by the case with the smallest number in the table, there will be a tendency of cases 1 or 2 to play this part. These cases tend to arise not only with genuine predictions, but also with the most trivial predictions, such as 'A will be conscious at time t_2' which, let us suppose, lies close to time t_1. This prediction belongs to case 1 and, if admitted, would imply that A does not possess free will according to any of the five definitions. However, A may reasonably say that such a prediction, which restricts his brain state at t_2 only in the most trivial way, is compatible with complete freedom of will for him. It will then be necessary to agree on some criterion of significance or non-triviality for predictions P, and to replace 'prediction' by 'significant prediction' in the Table and elsewhere. The choice of a criterion of significance is evidently an important problem, but we shall not consider it here.

Our conclusions are: (1) In all cases considered the objective definition *D1* free will is absent in a mechanistic universe. (2) If only predictions belonging to case 7 (a subclass of the class of secret prediction) are made with respect to an agent, and if this agent rejects the definition *D1*, then he may consider himself as having *D2*-free will, though other persons would regard him as not having any free will. (3) Definitions *D3*, *D4*, *D5* are introduced for purposes of analysis but are finally judged to be unsatisfactory.

D. A. EVANS
P. T. LANDSBERG
Department of Applied
Mathematics and Mathematical
Physics, University College
Cardiff

REFERENCES

Good, I. J. [1971]: 'Free Will and Speed of Computation', *British Journal for the Philosophy of Science*, **22**, pp. 48–9.

Landsberg, P. T. and Evans, D. A. [1970]: 'Free Will in a Mechanistic Universe', *British Journal for the Philosophy of Science*, **21**, pp. 343–58.

MacKay, D. M. [1965]: 'Information and Prediction in Human Sciences', in S. Dockx and P. Bernays (*eds.*): *Information and Prediction in Science*, pp. 255–69.

MacKay [1967]: *Freedom of Action in a Mechanistic Universe* (Eddington Memorial Lecture).

MacKay [1971]: 'Choice in a Mechanistic Universe: A Reply to Some Critics', *British Journal for the Philosophy of Science*, **22**, pp. 275–85.

Watkins J. W. N. [1971]: Freedom and Predictability: An Amendment to Mackay', *British Journal for the Philosophy of Science*, **22**, pp. 263–86.

I'M FREE BECAUSE I KNOW THAT I DON'T YET KNOW WHAT I'M GOING TO DO?

In 'Freedom and Predictability: An Amendment to MacKay' (Watkins [1971]) Watkins states that the claim MacKay is making in his freedom articles is that there can be no predictive scientific knowledge about some man, *A*, that would establish that his feeling of freedom is illusory (Watkins [1971], p. 272). Watkins emphatically denies that MacKay is attempting to demonstrate that man is free (Watkins [1971], pp. 264, 269). I think that it is fairly safe to say that Watkins has completely misunderstood MacKay's intent. In 'Choice in a Mechanistic Universe' (MacKay [1971], henceforth *CMU*) MacKay (quoting from his earlier work, *Freedom of Action in a Mechanistic Universe* (MacKay [1967], henceforth *FAMU*)) states: 'In order to *demonstrate human freedom* . . . (*sic*) it is unnecesary . . . (*sic*) to repudiate mechanistic brain theory' (*CMU*, p. 284, and *FAMU*, p. 37, my italics). He says in the same article: 'Our freedom in reality may be even greater than that which *I have been concerned to demonstrate.* My argument is that it cannot (logically cannot) be less' (*CMU*, p. 280, my italics.) MacKay goes on to say that this least freedom is the power that *A* has to determine his future actions (*CMU*, p. 283).

I will argue in this discussion that MacKay's attempted demonstration of human freedom fails on two counts. I will show first that MacKay is arguing as if the only determinists who claim that man is not free are those determinists who *also* claim that human actions are predictable; since this is not the case, what I take to be MacKay's claim that human actions are predictable and man is free (even if it were true) would in no way imply that man is free if the universe is (a-predictively) deterministic. Secondly, I will consider this claim that human actions are predictable and man is free, show that it is in fact the claim MacKay is making, and then show that it rests on the unjustifiable (or at least unjustified) assumption that what is indeterminate-for-*A* is necessarily determinable-by-*A*.

In 'Indeterminism in Quantum Physics and in Classical Physics' (Popper [1950]), an article which MacKay often refers to, Popper distinguishes quite clearly between two different 'doctrines of determinism': the scientific determinist claims that given a sufficiently powerful calculator, there is no future event which is not predictable in accordance with the methods of science (Popper [1950], p.

120); the metaphysical determinist claims 'that future events are, even before they happen, unalterably fixed, in exactly the same sense that past events are fixed' (Popper [1950], p. 121). Obviously, as Popper points out, the scientific determinist's claim is the stronger of the two; the scientific determinist is necessarily a metaphysical determinist, but the metaphysical determinist need not be a scientific determinist. Mackay appears to accept Popper's distinction; at least he says that what he calls scientific or physical determinism is the same as what Popper (and Watkins) call metaphysical determinism (*CMU*, p. 283), and I think that MacKay's metaphysical determinism is the same as Popper's (and Watkins') scientific determinism (see *CMU*, pp. 283–4). In an attempt to eliminate confusion I will refer to Popper's metaphysical and MacKay's physical or scientific as *a-predictive determinism* and to Popper's scientific and MacKay's metaphysical as *predictive determinism.*

I will argue later that if MacKay wants to demonstrate that man is free even if the universe is deterministic, he will have to confront the claim of the a-predictive determinists. This is something which as far as I know he has never done. When he claims, for instance, in 'On the Logical Indeterminacy of a Free Choice' (MacKay [1960], henceforth *LIFC*) that man is free even if the universe is deterministic, the variety of determinism he has in mind is clearly predictive determinism. He begins the article by presenting what he considers to be the usual form of the determinists' argument.

> If my physical brain-processes were wholely physically determined, and if my decisions could be inferred uniquely from my brain-processes, then a fully-informed observer of my brain-processes could know the outcome of my choices with certainty before I made them, and my impression of freedom in making these choices would therefore be an illusion, due to mere ignorance of the true state of affairs (*LIFC*, p. 31).

This formulation would be unacceptable to the a-predictive determinists since it includes the claim that a man's decisions can be inferred uniquely from his brain processes. The argument MacKay is examining, then, the predictive determinists' argument, is based (as MacKay's formulation of the argument indicates) on two 'assumptions':

(*PD* 1) if the universe is deterministic, then human actions are predictable;
(*PD* 2) if human actions are predictable, then man is not free.

MacKay's argument is that he can grant (for the sake of argument) that the universe is (predictively) deterministic and still demonstrate that man is free.

MacKay begins his demonstration by asking the question: 'How cogent, in fact, is the initial assumption that a fully mechanistic theory of our brain-processes would prove our actions inevitable, and our freedom an illusion?' (*FAMU*, p. 7). He then goes on to limit this question (which as it stands would be a challenge to both a-predictive and predictive determinists), asking us to 'suppose that all the relevant facts on the workings of your brain could be made available without disturbing it, to a computer system capable of predicting its [your brain's] future behaviour from these facts and the environmental forces acting on your nervous system' (*FAMU*, p. 8). The question in my mind, of course, is which of the two initial predictive determinist assumptions does MacKay mean to question. Or does he mean to question both? In other words, is MacKay going to try to show *PD* 1 to be false by demonstrating that the universe is deterministic and that human actions are not predictable? Is MacKay going to try to show *PD* 2 to be false by demonstrating that human actions are predictable and that man is free?

We will consider first the possibility that MacKay is questioning *PD* 1. At first glance, this might seem to be what he is doing, for he assumes (for the sake of argument) that the universe is deterministic, and he argues that human actions are never *unqualifiably* predictable. There are two reasons, however, why I think it is highly unlikely that MacKay is questioning *PD* 1. First, even if *PD* 1 is false, this does not imply that man is free. Secondly, unless *PD* 1 is true, there would be no way for MacKay, by examining the limits of predictability, to show (even if he demonstrates that man is free) that human freedom is compatible with a deterministic universe. MacKay, then, must be questioning *PD* 2. He is arguing, I think, that *B* (under certain conditions) can make true predictions about *A*'s future actions, and *thus human actions are predictable.* He is also arguing that since one of the conditions which must obtain in order for *B* to make true predictions is that *A* not know these predictions, *A*'s future actions (predicted by *B*) are not known to *A*, or as MacKay says, the actions are 'indeterminate-for-him' (*CMU*, p. 283), and *thus man is free.*

It should now be clear why MacKay's principal concern in his freedom articles is to show that there are situations which are at the same time both determinate and indeterminate, or in his words, that there are situations which manifest 'an inescapable element of relativity' (*CMU*, p. 276). He tries to establish this claim by arguing that if *B* (without telling *A*) predicts one of *A*'s future actions, it cannot be said of *A* that he is *ignorant of the true state of affairs* (that is, of the future state of affairs referred to in *B*'s true prediction), for if *A* knew *B*'s prediction he could then act in a way which would falsify it (see *LIFC*, pp. 35, 36, 39, and *CMU*, pp. 279, 281–3). MacKay concludes from this that there are *two* true descriptions of the state of affairs with regard to the future action of *A*. The true description for *B* is that *A* will do *X* at t_2; the true description for *A* is that his action at t_2 is indeterminate. Now I will assume (for the sake of argument) that MacKay is correct, that these two different descriptions of the state of affairs with regard to the future action of *A* at t_2 are both true. Is MacKay correct in then making the further claim that this relativity demonstrates that *A* (or man) is free?

That MacKay is actually making this claim that if *A*'s future actions are indeterminate for him, then he is free, is evident, I think, from the following passage from a conversation among *A*, *B*, and *LI* (the Laplacean Intelligence):

> *A*: So I was correct in believing I was free, even though my brain went through physically-determinate motions?
>
> *LI*: Yes. There is no need to refute what Thalberg calls 'the Deterministic Principle, that every event is caused' in order to recognise the objective reality of human freedom.
>
> *B*: But surely from *your* standpoint *A* wasn't free?
>
> *LI*: Careful! From my standpoint *A*'s *brain-and-body* had a determinate future course; but the very same evidence showed me that *A-as-a-cognitive-agent* faced an outcome *indeterminate-for-him* which it was in his power to determine. In this sense, from my standpoint also, *A* was free (*CMU*, p. 283).

MacKay seems to be saying here (through the mouth of the Laplacean Intelligence) that if a future outcome is indeterminate for A, then A has the power to determine that outcome and that since A has the power to determine the outcome, A is free. It is not at all clear to me what MacKay means by 'power to determine'. But if he means that A (as opposed to A's environment interacting with A's brain-and-body) has the power to determine whether he will do X or Y at t_2, then I would argue that MacKay has in no way shown (even if the description of the state of affairs true for A is different from the the description of the state of affairs true for B) that A has such power to determine. If MacKay means something else by 'power to determine', then I fail to see how his conclusion that man is free says anything more than that man feels free.

It may turn out of course that MacKay can show that indeterminate-for-A implies determinable-by-A. To do so would be, I suppose, a demonstration of human freedom. But it would also be a refutation of the claim that the universe is predictively deterministic. For as I suggested above, the predictive determinists' claim (that given a sufficiently powerful calculator there is no future event which is not predictable in accordance with the methods of science) is true only if the a-predictive determinists' claim (that future events are, even before they happen, unalterably fixed, in exactly the same sense that past events are fixed) is true. And this claim is surely inconsistent with the claim that man is free. As Watkins says: 'We must not expect too much from a discussion [MacKay's] which begins by conceding *metaphysical* [a-predictive] determinism to the opposition. If A's brain is as mechanical as clockwork, and if all his movements are causally predetermined by purely physical factors, then that by itself, it seems to me, sufficiently establishes that, *whatever* he does at t_2 and whether what he does is or is not scientifically predictable, any accompanying feeling of freedom he may have is illusory' (Watkins [1971], p. 269). For the time being, until Mackay refutes the a-predictive determinist's claim and shows that indeterminate-for-A implies determinable-by-A, we can say no more about his effort, it would seem, than that he has substituted for the old: 'I'm free because I feel free' the more cumbersome cry: 'I'm free because I know that I don't yet know what I'm going to do'.

JOHN MCDERMOTT
Eisenhower College

REFERENCES

MACKAY, D. M. [1960]: 'On the Logical Indeterminancy of a Free Choice', *Mind*, **69**, pp. 31–40.

MACKAY, D. M. [1967]: *Freedom of Action in a Mechanistic Universe* (Eddington Memorial Lecture).

MACKAY, D. M. [1971]: 'Choice in a Mechanistic Universe: A Reply to Some Critics', *British Journal for the Philosophy of Science*, **22**, pp. 275–85.

POPPER, K. R. [1950]: 'Indeterminism in Quantum Physics and in Classical Physics', *British Journal for the Philosophy of Science*, **1**, pp. 117–33, 173–95.

WATKINS, J. W. N. [1971]: 'Freedom and Predictability: An Amendment to MacKay', *British Journal for the Philosophy of Science*, **22**, pp. 263–75.

Brit. J. Phil. Sci. 23 (1972), 347–361 *Printed in Great Britain*

Review Article

LOUIS PASTEUR—A CASE AGAINST REDUCTIONIST HISTORIOGRAPHY*,[1]

Since scientists tend to publish and preserve records only of that work which they consider successful, there is even in the primary material of the historian of science, a heavy slant in the direction of a continuous success story. What I shall call *reductionist historiography*, by evaluating past scientific achievements relative to present science rather than relative to the contemporary alternatives to such achievements, further 'purifies' the material and further distorts the story. There often results a neglect of the degenerating phase of a successful research programme and a corresponding underestimation or misinterpretation of the theoretical ideas guiding the research. Also, scientific controversy is toned down, as if it were essentially unscientific, as if it were a result of external personal or political preferences illegitimately intruding on the internal life of pure science.

An interesting case is provided by historians' interpretations of the work of Louis Pasteur. Pasteur's work had a great impact on biological science and many of his ideas were highly controversial. Also, a study of the anti-reductionist tendencies in Pasteur's biological research is of special interest in view of the domination of modern biology by physical reductionism. However, some of the most interesting characteristics of Pasteur's work seem to have been missed by historians. The discussion of problems concerning the reduction of biology to physical science is in a rather confused and confusing state at the moment, and this is, I think, partly due to the strait-jacket current reductionist views put on the history of the development of science.

I shall first try to make clear what I shall mean by the terms 'reductionism' and 'antireductionism' in the methodology of biology. (*Physical reductionism*—the idea that biology is, in some sense, reducible to physics—is not to be confused with *historiographical reductionism*—the writing of history of science as a cumulative process issuing in our present body of knowledge. Indeed, as I point out *below*, page 348, one can perfectly well be a physical antireductionist and nevertheless write reductionist history. I hope this double use of the term 'reductionism' will not lead to confusion.)

It is important to distinguish *philosophical* from *methodological* reductionism/antireductionism. The philosophical position is concerned with the long run development, limited by 'conditions of possibility' ('*Bedingungen der Möglichkeit*' in the terminology of German idealism); while the methodological position is concerned with the strategy to be adopted in current research.

Philosophical *anti*reductionism is not inconsistent with methodological reductionism. Immanuel Kant, for example, saw definite limits to the analysis of living organisms by physical methods (such limits following from his general theory of the human mind), but at the same time, he urged that this analysis should be pursued as far as possible since only physical science gives truly scientific knowledge. The famous French physiologist Claude Bernard, contemporary of Pasteur, seems to have been, in this respect, a follower of Kant. His whole work in experimental science was aimed at chemical explanation of biological phenomena, but at the same time he saw definite limits to this kind of explanation. Like Kant he thought that explanation of the purposeful structure and development of living organisms as treated by morphology and embryology was beyond this limit (Bernard [1878–9], vol. 1, pp. 292 ff.). This is a good example of how a scientist may *accept* as the hard core of his research programme a heuristic principle, which he does not *believe* to be a general truth about the universe (cf. Watkins [1958], pp. 355–7).

---

* I want to thank especially Imre Lakatos and John Worrall, for useful and stimulating discussions and criticisms during my work with this paper.

[1] Review of R. Dubos [1951]: *Louis Pasteur, Free Lance of Science*, London: Victor Gollancz, pp. 418, and F. Dagognet [1967]: *Méthodes et Doctrines dans l'Oeuvre de Pasteur*, Paris: Presses Universitaires de France, pp. 272.

On the other hand many modern philosophical reductionists would allow that antireductionist (or at least non-reductionist) research programmes may have great heuristic value—in preparing the way for future reductions (cf. Schaffner [1967]). Thus philosophical reductionism and methodological antireductionism are consistent too—Pasteur can possibly be viewed as having adopted this combination of positions. Because of his strenuous attempts to *disprove* by experiments the reductionist rivals to his own theories, Pasteur must be classified as a methodological *anti*-reductionist rather than just a *non*-reductionist. But it is more difficult to locate him on the philosophical dimension. In contrast to Bernard, for example, Pasteur showed little interest in philosophical problems. His speculations and attempts to do experiments on the role of 'asymmetric forces' in the production of biological material (see p. 355, *below*) could be taken to express an underlying reductionist philosophy. But by themselves these ideas were certainly intended as empirically testable hypotheses and not as stating definitive philosophical principles. Also his works contain many statements which accord well with Bernard's antireductionist philosophical views. (Of course, if one does not think that any definitive delimitation of possible scientific developments is feasible, then one will simply adopt one of the two *methodological* positions.)

It is also important to see that antireductionism in biology (with respect to a reduction of biology to physical science) is compatible with historiographical reductionism. In fact the two authors with whom I shall be mainly concerned in the following, both seem to believe in the ultimate irreducibility of biology to physical science. The antireductionist argument often presupposes a particular physical theory, or part of such a theory, claiming that there are certain biological phenomena which this theory cannot encompass (cf. Roll-Hansen [1969]). Thus reductionist historiography, evaluating past science relative to present theories, is consistent with both positions on the reductionist/antireductionist controversy in biology.

René Dubos in *Louis Pasteur, Free Lance of Science* and Francois Dagonet in *Méthodes et Doctrines dans l'Oeuvre de Pasteur*, analyse Pasteur's work from very different points of view, but both concur in their reverence for the master, and both tend to disguise his failures.

Dubos, himself an active and prominent biologist, emphasizes Pasteur's discovery of 'hard facts' and makes little attempt at philosophical sophistication.

Pasteur's ability to provide experimental support for his theories was certainly a major factor in his success, but we must also explain for his rather dogmatic insistence on certain theories in the face of adverse evidence or experimental failure. Dubos's story does little towards providing such an explanation.

One of Dubos's soothing conclusions is that there was 'no real justification' for the famous controversy over fermentation[1] since there was agreement on the established facts. The conflicting theories 'were not incompatible. . . . Minor adjustments would have sufficed to compose the differences between the proponents of the two theories' (Dubos, p. 197). Dubos's statement that 'the relationship between optical activity and life still appears to be as close as suspected by Pasteur' (p. 110) shows the same dubious striving for reconciliation. The strength of Dubos's account lies in his specific biological insights and in his sympathy for Pasteur's more speculative views on the relation of life and matter.

In Dagognet's analysis the ideal of harmonious scientific development towards one definitive goal appears in the guise of the doctrine that any scientific theory is *inevitably* a mixture of truth and falsity. It is the fate even of the best scientific theories that they are later found to have many false consequences. Nevertheless, Pasteur's theories contain an important kernel of uncontested truth. Pasteur is not to be blamed for the false consequences of his theory since they were, apparently, dictated by the circumstances:

> 'There is no risk of our analysis becoming irreverent, since it also stresses the ineluctable character of the pasteurian interpretation. To us, the falsity that is mixed with the truth, seems under the circumstances inevitable.' (Dagognet, p. 44.)

Dagognet ascends to very abstract heights when defending Pasteur's theories in stereochemistry and fermentation. One conclusion is that in the 'philosophical' contest with Berthelot's reductionist programme for organic chemistry and fermentation (see *below*, p. 358) the Pasteurian can still look to the future with confidence (Dagognet, p. 81 ff.). One gets an uneasy feeling that a level of abstraction has been reached where the results of empirical science are no longer relevant. But Pasteur and Berthelot both saw the dispute primarily as one to be experimentally decided.

A most interesting part of Dagognet's account is the demonstration of how the implicit doctrine which he calls 'vegetalism' connects the different phases of Pasteur's work. This doctrine, which aims 'to separate the living and mineral domains, to create contrast' (p. 8), is followed from its sources in crystallography and physical optics through the modifications and additions wrought by its successive applications to problems of stereochemistry, fermentation, pathology etc. While Dubos claims that Pasteur's germ theory is 'merely a body of factual observations which allows a series of practical operations' and that his interpretation of the results of the experiments on spontaneous generation are 'the permanent rock on which were built whole sections of biological sciences' (Dubos, p. 187), Dagognet stresses Pasteur's ability to reinterpret old recalcitrant facts and marshal new ones to support his own theories: 'Pasteur did not despair at reversing the sense of the most unchallengeable facts' (Dagognet, p. 113).

Although Dubos and Dagognet thus have very different views, each is a reductionist in the historiographical sense, i.e. tends to evaluate Pasteur's results in terms of truth and falsity, where what is true and what false is defined by present scientific knowledge. Dubos tries to show how some of the most important currently accepted facts were discovered by Pasteur. Dagognet argues that Pasteur's theories contained as much truth as one could possibly expect and that our most basic theoretical conceptions of the relation between the living and the non-living have hardly advanced since Pasteur.

---

[1] Cf. below, p. 356.

I would claim, however, that an approach which rated the merit of theories mainly in relation to their contemporary alternatives, could more easily account for the changing fortunes of Pasteur's research. Such an approach would be less committed to defending the greatness of the master by asserting the eternal truth of his results. Though the understanding of old theories can hardly avoid relying on some knowledge of present science, these theories can be *evaluated* relative to their contemporary alternatives. A more relational historiography may lessen—though not eliminate—the need for continuous rewriting of the history of science. Also evaluations of past science will then be more comparable to evaluations of *present* science, where one cannot profit from the wisdom of hind-sight.

In the following I try to give such a non-reductionist account of two central themes in Pasteur's scientific development and contrast this account with those of Dubos and Dogognet.

(*a*) *Tartaric acids and stereochemistry*

Pasteur began his scientific career in chemistry working on the problem of correlating certain crystallographic and optical properties of substances with their molecular constitution. He was stimulated by the discovery, made by the German chemist Mitscherlich, that an aqueous solution of tartaric acid had the power of rotating the plane of polarisation of light, while paratartaric (also called racemic) acid was quite inactive in this respect, though the two acids were identical in all other known chemical, crystallographic and physical properties. Pasteur worked from the idea that there must be some difference in the constitution of the two acids responsible for the difference in their optical activity.

(The statement about ordinary tartaric and paratartaric acid having identical properties needs some qualification. From the beginning the two acids were clearly recognised as different substances. The paratartaric acid had been discovered around 1820 among the byproducts of wine making together with the ordinary tartaric acid. But a few years later it was no longer to be obtained from the original source, without anybody being able to explain why. The paratartaric acid and its strange behaviour caused considerable interest among the chemists of the time.)

In 1848 Pasteur made his first great discovery, namely that the optically inactive paratartaric acid could be separated into two optically active fractions. One was identical with the naturally occurring tartaric acid and rotated the plane of polarisation to the right, while the other fraction rotated the plane to the left by an equal amount.

This first discovery provides a nice illustration of Pasteur's forceful way of applying a hypothetico-deductive procedure. On the hypothesis that optical activity is correlated with asymmetry of the crystals, he checked a whole series of salts of the paratartaric and tartaric acids, and was very pleased to find that the salts of tartaric acid distinguished themselves by forming asymmetric crystals (crystals that could not be superimposed on their own mirror-image), while all the paratartrates formed symmetric crystals. This was an observation which had evaded Mitscherlich and other earlier investigators.

But when Pasteur came to the sodium ammonium salt of paratartaric acid, his hypothesis apparently broke down: the crystals were asymmetric. But he also noticed that they were of two forms, one being the mirror-image of the other. True to his idea about correlation between crystal form and optical activity, he separated the two crystal forms by hand and tested their optical activity separately. The apparent refutation of his hypothesis turned into a surprising confirmation: the inactive paratartaric acid could be split into two fractions, one being identical with the naturally occurring tartaric acid and the other being its optical antipode.

This discovery brought Pasteur into close contact with Biot who had been investigating the optical activity of diverse organic materials and who came to have an important influence on Pasteur's speculations about the connection between molecular asymmetry, optical activity and life. Pasteur maintained the vegetalist position that all naturally occurring asymmetric molecules originated from living beings and could not be synthesised in the laboratory, at least not by the methods of his contemporary chemistry. At first Pasteur's theories in stereochemistry were very successful, and he was able to rebut opponents by incorporating their discoveries into his own system and at the same time predict new facts. But after about 1855, when Pasteur switched his main effort to fermentation and other problems of microbiology, his research programme in stereochemistry soon showed signs of faltering, and by 1880 it was reduced to an exercise in devising *ad hoc* hypotheses to protect it from new facts.

In two lectures given to the Chemical Society of Paris in 1860 Pasteur sums up his work on 'Molecular asymmetry of naturally occurring organic compounds'. He formulates two basic principles.

1. Chemical compounds generally come in two forms, with symmetric and asymmetric molecular arrangements respectively. To every asymmetric compound there is a corresponding symmetric one.

> '. . . the groups of elementary atoms, from which complex matter is constituted, may assume two different conditions which correspond to the two general types to which all material bodies can be referred. The form of the group is either identical with its reflected image or it is enantimorphous with it.' (*Oeuvres*, vol. 1, p. 337.)

2. Only the products of living organisms are asymmetric.

> "All artificial bodies and all minerals have superimposable images. Opposed to these are most organic substances (I might even say all, if I meant only those which play an important role in plant and animal life), all the products that are essential to life, are asymmetric and have the kind of asymmetry in which the mirror-image cannot be superimposed on the object.' (*Oeuvres*, vol. 1, p. 331.)

The analogy between asymmetric as against symmetric molecules on one hand and life *contra* non-life on the other, is the basic theme that runs through these lectures. Pasteur's purpose is to separate 'the chemistry of non-living nature from the chemistry of living nature' (*Oeuvres*, vol. 1, p. 343). Toward the end he expresses his great faith in this as a programme for future research:

> 'You will in the course of these lectures have understood why I have called them "On the molecular asymmetry of natural organic compounds". I have in fact set up a theory of molecular asymmetry, one of the most elevated and completely unforeseen chapters of science, which opens up new distant, but certain, horizons for physiology.' (*Oeuvres*, vol. 1, p. 343.)

In accordance with these principles, and his experimental results, Pasteur thought there were in the simple case[1] three stereochemical isomers—as illustrated by the tartaric acids:

'Active bodies: molecular groups	Right. Example: ordinary tartaric acid. Left. Example: left tartaric acid.
Inactive bodies: molecular groups	Neither right nor left *by nature.* Example: inactive tartaric acid. Neither right nor left *by compensation.* Example: paratartaric acid.'

(*Oeuvres*, vol. 1, p. 346)

By 1860 Pasteur felt quite sure that his theory, including this schema was correct. For example, he had in 1853 predicted and successfully produced mesotartaric acid, which he took to be the inactive symmetric isomer (inactive 'by nature'). Among the products of this experiment he also unexpectedly found the paratartaric acid (*Oeuvres*, vol. 1, pp. 257 ff.).[2]

Following this classification, Pasteur imputed the symmetric molecular form to many inactive substances which, according to present chemistry, are just a mixture of left (l-) and right (d-) forms, and thus fit none of Pasteur's four categories.

For example, Dessaignes in 1850 announced the conversion of maleic and fumaric acids, earlier found to be optically inactive, into aspartic acid, earlier found to be optically active. But Dessaignes had not directly shown his product to be optically active. Pasteur immediately proceeded to test the optical activity of Dessaignes's aspartic acid. As Pasteur predicted from his own theory, Dessaignes's aspartic acid was inactive. Thus Pasteur could interpret the experiment as an important confirmation of his own theory (*Oeuvres*, vol. 1, pp. 155 ff.).

---

[1] In the simplest case (of a molecule with one asymmetric centre) one can, according to Pasteur, have three isomers, the d- and l-forms, and the symmetric inactive form. The paratartaric acid contains two asymmetric centres, which in this special case exactly compensate each other. But in general two asymmetric centres will give four kinds of asymmetric molecules, three asymmetric centres eight kinds etc. Pasteur further speculated that with many (more than two) asymmetric centres in one molecule, crystallisation would be impossible since there would not be enough possible forms of crystallisation. This, he suggested, might be the explanation for albumins not being crystallisable (*Oeuvres*, vol. 1, p. 380).

[2] For this synthesis of paratartaric acid (and some earlier work pertaining to its origin) Pasteur was awarded a prize set up by The Pharmaceutical Society (see report by Mr Grassi reprinted in Pasteur's *Oeuvres*, vol. 1, pp. 468 ff.). Dagognet's use of this work to illustrate Pasteur's fidelity to the ideal of pure science is scarcely convincing. He writes: 'Pasteur cherishes disinterested research, he is dedicated to the solution of apparently useless problems' (Dagognet, p. 59). Through his whole career Pasteur took up problems of general public interest. His work on spontaneous generation and rabies, for example, were well planned and extremely successful propagandistic moves (this by no means detracts from their scientific value, but rather enhances it).

With little hesitation Pasteur took Dessaignes's aspartic acid (which he further converted into malic acid) to be the symmetric rather than the racemic (inactive 'by compensation') form. His argument, from the modern viewpoint, seems curiously unconvincing:

> 'It might be asked whether the new malic acid is not the paratartrate of this group, that is, the compound of the right and left malic acids. This is highly improbable since in that case not only would one active substance have been produced from an inactive substance, but two, one dextro- and one laevorotatory.' (*Oeuvres*, vol. 1, p. 349.)

In considering Pasteur's dismissal of the racemic alternative it is essential to understand that he thought of the 'racemic form' as a compound molecule and not as the mixture of different kinds of molecules, which is called 'racemic' in modern chemistry. Pasteur thought of the d- and l- acids as being produced from the racemic acid by a splitting of the molecule and not merely by the separation of two different kinds of molecules. In 1861 he writes:

> 'Nothing is proved with more assurance than the molecular constitution of paratartaric acid. The right and the left tartaric acids combine, equivalent by equivalent, giving off heat and forming by their union the paratartaric acid.' (*Oeuvres*, vol. 1, p. 349.)

Dubos does not distinguish clearly between Pasteur's conception of the racemic form and the modern conception, and tends to read later theoretical developments into Pasteur's work (pp. 109 f.). Dagognet is not very clear either (pp. 49 ff.). This lack of a clear interpretation of Pasteur's stereochemical theory is unfortunate since it makes his stubborn rearguard actions during the seventies and eighties seem more irrational than is necessary. Neither Dubos nor Dagognet gives much space to this methodologically interesting phase of Pasteur's research programme in stereochemistry.

When Perkin and Duppa in 1860 announced that they had made tartaric acid from succinic, Pasteur again showed that the product was optically inactive, and thus not identical with the naturally occurring tartaric acid, as the original announcement seemed to imply. But Pasteur did find it to contain the racemic acid which can be split into d- and l- forms. To protect his stereochemical principles he tentatively argued that the succinic acid from which the synthesis started was a natural product and could well itself be the racemic form rather than the symmetric. One would expect succinic acid to occur in the same four forms as tartaric, says Pasteur (*Oeuvres*, vol. 1, pp. 345 ff.). On the modern theory succinic acid cannot have optically active isomers.

Later when succinic acid had been synthesised from inorganic materials alone, Pasteur admits that his stereochemical principles have taken on 'a new character', but he still holds them to be valid (*Oeuvres*, vol. 1, p. 358).

In a 1875 paper entitled 'On a distinction between natural organic products and artificial organic products', Pasteur defended his 1860 stereochemical ideas by insisting that to produce racemic acid is not to produce an active body.

This is not achieved until the racemic acid is split into d- and l- forms (*Oeuvres*, vol. 1, pp. 364 f.). This defence clearly depends on the racemic acid being a specific compound molecule and not just a mixture of d- and l- molecules. It is thus misleading of Dubos to attempt to explain and defend Pasteur's point of view with the argument that even today one can by purely chemical methods only produce mixtures of d- and l- compounds and that some kind of asymmetric input is needed to separate them (Dubos, p. 110)—like picking them out by hand or by help of a programmed machine. For Pasteur the important thing is that 'asymmetric forces' are needed to split the molecule of the racemic acid.

In the same 1875 paper Pasteur tries to justify his theory by pointing to its predictive successes. He mentions the case of Dessaignes and that of Perkin and Duppa, described above, as well as a case concerning mannitol (*Oeuvres*, vol. 1, pp. 364 f.).

But the barren *ad hoc* character of Pasteur's modifications of his theory become increasingly clear, and it becomes more and more difficult for him to defend the idea that special organic or other asymmetric forces are necessary for the production of asymmetric molecules. In 1883 Pasteur in a new talk to The Chemical Society of Paris on 'Molecular asymmetry', repeats his main principles of 1860. But by this time most organic chemists apparently were finding it useless to explain the splitting of the sodium ammonium salt of paratartaric acid into d- and l- tartaric acids by the 'presence of organic particles in the solution or on the surface of the vessels of crystallization' (*Oeuvres*, vol. 1, p. 381).

In clarification of Pasteur's position it must be added that he did not think it absolutely impossible to produce asymmetric molecules artificially. But he thought that completely new methods would have to be introduced, methods using 'asymmetric forces'. In the fifties especially Pasteur had done many experiments trying to induce asymmetry in organic compounds by asymmetric influences derived mainly from optical and magnetic sources. As he recollects in the abovementioned lecture of 1883:

> 'Already at Strassbourg, I had powerful magnets constructed by Rhumkorff; in Lille, I made recourse to revolving motion, provoked by clockwork. I wanted to try to grow a plant from its germination under the influences of sun rays inversed by a mirror governed by a heliostat.' (*Oeuvres*, vol. 1, p. 376).

He still believed in the research programme set forth in his 1860 lecture: the compounds characteristic of life originate under the influence of asymmetric forces, but the chemists use no asymmetric forces and therefore produce no asymmetric molecules. We have to look for completely new methods to be able to imitate nature, says Pasteur. For example in some notes from 1870–1, sketching possible experiments, he speculated that if the crystallisation of sodium ammonium paratartrate took place in a completely dark room, you might not get a splitting into d- and l- tartaric acids (*Oeuvres*, vol. 7, p. 21). He complained in 1874 that to synthetise asymmetric molecules is a much more difficult, profound and important problem than many people thought (*Oeuvres*, vol. 1, p. 358).

Pasteur's speculative views on the physical basis of life have some of the same romantic flavour that Gunther Stent ascribes to the molecular biology of the 1940s and 1950s, the 'heroic' period when many people expected that fundamental changes in physics would be necessary to explain biological phenomena, that molecular biology would inspire a new revolution in physics (Stent [1968]).

What I called the two principles of Pasteur's stereochemistry, that chemical compounds may have two molecular forms, symmetric and asymmetric, and that all asymmetric molecules are products of living organisms, formed the 'hard core' of his research programme. These principles apparently made him unresponsive to the idea of the asymmetric carbon atom[1] proposed around 1870 by J. H. van't Hoff, J. A. LeBel and J. Wisclicenus, and others. This new idea, which was closely linked with the contemporary development of valence theory and structural organic chemistry implied that in the simple case there were only two optical isomers for each compound which included one asymmetric carbon atom, the d- and l-form. Specifically there was no inactive isomer with a symmetric molecular structure. So, if certain compounds were produced at all, the molecules were necessarily asymmetric, and Pasteur's asymmetric forces lost much of their *raison d'etre.*

In the new theory there was accordingly little room for Pasteur's fundamental division between the chemistry of the living and the chemistry of the non-living which, as we have seen, was a basic element of his theory. The basis of Pasteur's antireductionist research programme in stereochemistry was destroyed by the introduction of a new and more successful theory which abolished the one side of Pasteur's dualism: the symmetric inactive isomer of the optically active compounds simply did not exist. Pasteur could hardly have retreated to the position that the inactive compound is a mixture of equal amounts of d- and l- molecules, and that this mixture can only be separated by some kind of asymmetric force, as suggested by Dubos (p. 110) without giving up at least the first of the two stereochemical principles listed above.[2] And if the first principle is given up the second one makes little sense.

The critics, against whom Pasteur fought around 1880, built their theories around the new idea of the asymmetric carbon atom, and this was to be the basis for the development of stereochemistry in the following decades. One could argue, however, that Pasteur's reluctance to give up his basic ideas in stereochemistry even after many fruitless years, was justified, or at least made explicable by the success of closely related ideas in microbiology.

Pasteur's very successful later research in microbiology was closely linked with his early work in stereochemistry. Pasteur's vegetalism, separating the living from the non-living in a fundamental way, was expressed in stereochemistry as the principle of the organic origin of asymmetric molecules, and in microbiology as the germ-theory—microscopically small living organisms were the causes of such processes as fermentation, putrefaction, disease, etc., which before Pasteur were considered to be purely chemical.

---

[1] When a carbon atom with four bonds, protruding like the corners of a tetrahedron combines with four different atoms or other radicals, one gets a tetrahedral molecule which is not superimposable on its mirror-image.

[2] Dubos is stretching the material excessively when he declares that with his discovery of mesotartaric acid, Pasteur 'had postulated and demonstrated the existence of a new type of optically inactive molecules' namely the kind of molecule where the effects of two asymmetric centres exactly cancel each other. As I have shown this kind of molecule is nearest in Pasteur's conception to the racemic acid molecule which he calls 'symmetric by compensation'. The mesotartaric acid he considered to be a simple case of a molecule 'symmetric by nature' (Dubos, pp. 109 ff.).

(*b*) *Fermentation*

In 1849, at Bliot's suggestion, Pasteur started work on the optically active amyl alcohol, a byproduct of ordinary alcohol fermentation. Gradually he was led on through the problems of fermentation into microbiology. But his famous researches on the fermentation of wine and beer etc. were not seriously underway till around 1855. He then took up a new position in Lille where one of his duties was to try to apply science to the problems of industry.

The dominant theories of fermentation at that time were mechanistic—or rather chemistic. The most important were Berzelius's theory that the ferment acted as a catalyst—like platinum on hydrogen peroxide, and Liebig's theory of fermentation as a sort of decay: the decay of albuminous organic matter brought about a splitting of the sugar molecule by an association with the sugar molecule transferring the destructive forces. Liebig's theory was the more widely accepted.

Pasteur soon joined the vitalist opposition which claimed that living organisms were essential to fermentation. Schwann, Caingnard de la Tour, Kützing and others had shown the presence of microscopic living organisms in fermentation processes. But their views gained little acceptance and were derided by Liebig and Berzelius. In 1857, in a paper on lactic acid fermentation, Pasteur formulated his germ theory of fermentation: fermentation is caused by the growth of germs and cannot take place without their vital activity (*Oeuvres*, vol. 2, pp. 3 ff.).

The concept of the germ was to be the core of Pasteur's very successful research into microbiology, where he defeated reductionist doctrines of spontaneous generation (namely that under appropriate conditions life will arise spontaneously from non-living matter) and the reductionist chemical theories of fermentation. Both these doctrines have, however, been successfully revived. The chemical interpretation of fermentation is now part of elementary biochemistry, and heterogenesis is at the core of thriving research into the origin of life and the possibility of creating it again.

But the new reductionist programmes are very different from those which Pasteur defeated. The new versions are essentially influenced by his work. Both his experimental procedures and his theoretical ideas are to a large extent built into them. It is only on a very abstract level that reductionism and antireductionism can be seen running through history as quite separate and independent scientific alternatives. But this does not necessarily imply that the conflicts between reductionist and antireductionist programmes are unimportant.

In his lactic acid paper Pasteur underlines the hypothetical character of his germ theory, that it goes far beyond the facts known at the time. But he argues that this is the case in all good creative research. His theory is justified by its success compared to the rival theories—both Berzelius's and Liebig's theories are 'contradicted by experience' (*Oeuvres*, vol. 2, p. 13). The positive evidence contained in this first paper lies mainly in the demonstration that a micro-organism is involved in lactic acid fermentation, though this micro-organism is much smaller than the yeasts involved in ordinary alcoholic fermentation.

In another paper, of 1858, Pasteur goes on to discuss some of the most important evidence of his opponents. Liebig argued that since yeast cannot grow on sugar and water alone, the fermentation which takes place when yeast is added to pure sugar-water cannot be caused by the growth of the yeast. Furthermore the acceleration of fermentation caused by adding 'albuminous material' to the sugar-water-yeast mixture, seemed to support Liebig's theory. Pasteur retorted that in the case of yeast in pure sugar-water the yeast is both growing and disintegrating. Growing yeast feeds on disintegrating yeast. In the second case the albuminous material serves as nutrient for the yeast, which therefore grows better and thus causes more fermentation.

Pasteur also developed new experiments to demonstrate that 'the facts on which Mr Liebig supports himself, do not have the exactness which he supposes'. Pasteur's seeding experiments are especially important. He showed that approximately the same amount of sugar that is caused to ferment by adding a certain quantity of yeast to pure sugar-water, will be caused to ferment when sugar-water supplemented by extract of the same quantity of yeast is seeded by a very small amount of living yeast cells. This also, says Pasteur, explains the previously puzzling fact that yeast which is boiled or dried at 100°C does not lose its fermentative power, although its action is delayed (*Oeuvres*, vol. 2, p. 19). Pasteur was soon able to improve his seeding-experiment by making the yeast grow in a solution containing no yeast extract, but only sugar, ammonium tartrate and ashes of yeast. In this case the starting solution contained no complex organic substances which could break down but only simple substances serving as food for the yeast. Apparently, Pasteur had eliminated the albuminous substances which, according to Liebig, caused the fermentation by their breakdown.

The Pasteur camp came to consider this as a crucial experiment. (See for example: Duclaux [1896], p. 75.) In a critique of Liebig's views in 1871 Pasteur describes this experiment as 'the touchstone of truth or error' (*Oeuvres*, vol. 2, p. 361).

Liebig, however, was not convinced. While hailing Pasteur's seeding experiment as a very important advance in plant physiology and admitting that the growth of yeast was very closely related to alcoholic fermentation, he could not accept Pasteur's vitalist ideas. He felt that Pasteur's work was irrelevant to the problems he wanted to solve:

> 'Mr. Pasteur has not concerned himself with what I have tried to explain, namely the decomposition of the fermentable substance in the presence of the yeast-cell, and in affirming that a vital act is the cause of the fermentation, he substitutes a fact for an explanation, a fact which is itself in need of explanation.' (Liebig [1871], p. 6.)

Pasteur on the other hand complains about the *ad hoc* character of Liebig's modifications to the decomposition theory, and about Liebig's inability to produce any new and enlightening experiments.

By 1871 Pasteur is convinced that the hypothesis, which he tentatively, but quite confidently, proposed in 1857, actually expresses the truth:

> 'I believe I have demonstrated by peremptory tests that a fermentable material never undergoes fermentation without there being a continuous exchange with living cells which grow or multiply while assimilating a part of the fermentable material itself'. (*Oeuvres*, vol. 2, p. 362.)

Generally Pasteur seems to have believed that scientific hypotheses, by means of successful testing, can be established as definitive factual truths at least in some cases:

> '. . . this marvellous experimental method about which we can truly say, not that it is sufficient for every purpose, but that it rarely leads astray, and then only those who do not use it well. It eliminates certain facts, brings forth others, interrogates nature, compels it to answer and rests only when the mind is fully satisfied. The charm of our studies, the enchantment of science, is that, everywhere and always we can give the justification of our principles and the proof of our discoveries. . . . The aim of the true experimental method is proof that is not open to any objection.' (*Oeuvres*, vol. 7, p. 334.)

This verificationist attitude to science is certainly part of Pasteur's rather dogmatic evaluations both of his own theories and their rivals, as is exemplified both in his stereochemistry and in his theories about fermentation.

After ferments like the one that inverts the rotatory power of sucrose by splitting the molecule into two, had been isolated as chemical substances, some people thought that similar chemical substances, might be responsible for alcoholic fermentation. The ferments were thought to be produced by the living yeast, but once produced they did not depend on the living cell for their activity. This new idea was seen by the chemists as a means of modifying the old chemical doctrine of Berzelius and Liebig.

Though Pasteur is reported to have made some attempts himself at isolating what we would today call the enzymes of alcoholic fermentation (Dubos, p. 201), he was negative in his evaluation of the new hypothesis, at least in the chemists' version. According to Pasteur processes like the inversion of sugar are very different from alcoholic fermentation. For example, the same inversion will be produced by any acid. It is only by the semantical trick of calling the substance that inverts sugar a 'ferment' that Berthelot is able to see any conflict between Pasteur's vegetalist theory of fermentation and the facts, says Pasteur (*Oeuvres*, vol. 2, p. 127).

In his *Studies on Fermentation* (from 1876 on) Pasteur insists on distinguishing between two fundamentally different orders of fermentation, caused by 'soluble' and 'insoluble' ferments respectively. The insoluble ferment is essentially bound to the living cell, while the soluble ferment can be separated from and act independently of the living cell. Pasteur points to the fact that while alcoholic fermentation is inhibited by high pressures of oxygen, hydrocyanic acid, salts of mercury, ether, etc., this does not affect the activity of soluble ferments like the one that inverts sugar (*Oeuvres* vol. 5, p. 252 f.). But he does not produce any important novel facts, like his earlier seeding-experiment, and the modification of his theory takes on a barren *ad hoc* character.

Berthelot formulated a clearly reductionist aim for the enzyme theory of fermentation, claiming that the facts generally supported his view. He wanted to:

> '. . . reduce the action of ferments to conditions which are purely chemical, . . . this is in effect what has been successively realized for nearly all fermentations, as is proven by the history of the glucosidic fermentation of the starch in the germinating barley, of the fermentation of fats in the intestine, of amygdaline in almonds, of sucrose which is inverted under the influence of yeast, of urea in urine etc. etc. Only two or three cases still remain obscure.' (Berthelot [1878], p. 587.)

Pasteur rejects this programme in a very polemical reply. But he also says that the eventual isolation of a soluble ferment capable of causing alcoholic fermentation, need not conflict with his views, as long as one accepts that 'the presence of micro-organisms is an absolute condition for fermentations properly called', 'that these organisms are not of spontaneous origin', and that in the absence of oxygen, any organism which can live under these circumstances, will produce alcohol (*Oeuvres*, vol. 2, p. 593).

At this point it seems that hardly anything but a full chemical synthesis of the alcoholic ferment could make Pasteur reconsider his basic doctrine of fermentation. So when Buchner succeeded in separating alcoholic fermentation from living yeast in 1897, two years after Pasteur's death, this was an important success for the new reductionist programme, but by itself insufficient to overthrow Pasteur's theory.

It seems that the most fruitful period of Pasteur's research on fermentation was the late fifties. Through the sixties and into the seventies he continued to refine his experimental procedures in the breaks between other work, but few important ideas arose pertaining to the controversy with the chemists. And by the time of the 1878 controversy with Berthelot, Pasteur's programme had run out of steam, the chemists were taking the lead again.

In fermentation as in stereochemistry Dubos and Dagognet overestimate the permanency of Pasteur's victories. They seem reluctant to acknowledge that after a very successful initial period Pasteur's research programmes stagnated. It makes little sense to say with Dagognet that the final outcome of the controversy between Berthelot and Pasteur over the chemistry of life, was that even 'Berthelot himself . . . admitted his defeat' and then substantiate this claim with a quotation from Berthelot *anno* 1860 (Dagognet, p. 80).

The problem of spontaneous generation was not at first very important for Pasteur's microbiological views. But as his germ theory developed with ideas about specific micro-organisms for each kind of fermentation, with ideas about the widespread presence of germs in the atmosphere etc., the non-existence of spontaneous generation became an important test for his views. If living organisms could arise spontaneously, for example in fermenting media, his vitalist theory would be seriously undermined. His mentors Biot and Dumas tried in vain to restrain Pasteur from entering the controversy over spontaneous generation, which they, according to Dubos, 'considered too complex for experimental approach' (p. 43).

Pasteur's eminent abilities both as an experimenter and a rhetorician are well illustrated in the debate over spontaneous generation. His famous public lecture of 1864, for example, is an impressive and simple display of the principles of his doctrine, illustrated by clarifying explanations of familiar phenomena and striking experiments (*Oeuvres*, vol. 2, pp. 332 ff.).

Pasteur complains how easy it is for his opponents by faulty technique to produce apparent instances of spontaneous generation, and how much harder it is to show in each case exactly where in the course of the experiment the germ has been unknowingly introduced. The germ theory supported by new techniques concerning sterilisation, handling of pure stains of micro-organisms etc. proved a very powerful instrument for detecting the mistakes of his opponents.

The lecture on spontaneous generation also illustrates Pasteur's conception of the external relationships of science. While religion, philosophy, politics etc. are irrelevant to purely scientific problems, the scientific facts must be respected in all aspects of human life. Therefore the correct solution of scientific problems may have very important implications for other fields. In this lecture Pasteur sees himself as defending religious and moral orthodoxy against atheism and materialism by demonstrating that the evidence claimed to support the doctrine of spontaneous generation is built on experimental mistakes.

> 'What a conquest for materialism if it could claim to be based on the proven fact of matter organizing itself into living beings, . . . Why then the idea of a creator God?' (pp. 332 f.).

Treating the scientific problem: 'Can matter organize itself spontaneously? . . . Can living beings come into the . . . world without parents, without an egg?' (p. 328); Pasteur assures his listeners of a negative answer. However, his explicit conclusion only states that 'in no circumstances known at present', can one affirm spontaneous generation (p. 346).

I have tried to show both the strength and the weakness of Pasteur's anti-reductionist research programme by describing a few examples of successes and failures from his research in stereochemistry and fermentation. I have not gone into his research in pathology, epidemology, etc., which occupied most of his later years, and in which he was probably even more unquestionably successful. As mentioned above, Pasteur's research programme in fermentation had stagnated by 1880 as his programme in stereochemistry had started to decline two decades earlier. In both cases the decline came at about the same time as Pasteur moved his main effort into new areas.

Why did Pasteur's ideas stop being productive in these areas? Did lack of success force him to shift fields, or did the theories falter because they lost the active support of Pasteur?

Pasteur's change from stereochemistry to fermentation around 1855 was indeed preceded by a period of extensive and unsuccessful experimenting with asymmetric forces. But it is clear from Pasteur's writings that he himself never came to think of his stereochemical research programme as exhausted. He repeatedly expressed his regret that he did not have time to take up again the fundamental problems concerning the relation between life and matter. And he continued to believe that asymmetric forces and the difference between symmetric and asymmetric molecules was a main key to these problems. His results in more practical fields were mere corrolaries to the basic doctrines about life and matter. In the case of fermentation no similar series of unsuccessful experiments seems to have taken place though this is difficult to confirm because of Pasteur's habit of not telling even his closest associates the idea behind the experiments they were conducting (Dubos, p. 59).

This problem of whether Pasteur's theories would have been more successful had he had more time for them is difficult but interesting. Its study might help give some substance to the present discussions about scientific objectivity. For example to the question: to what extent can the theoretical development of science be influenced, or even steered, by conscious decisions of science policy, or more subtle forms of pressure?

Apart from this problem, I think the above case-history shows clearly enough how Pasteur succeeded in putting the study of some natural phenomena of classical interest to science on a new biological footing. Phenomena like fermentation, putrefaction, disease etc. were, at the beginning of Pasteur's career, considered by most scientists as chemical and mechanical. Pasteur formulated distinctly biological theories for these phenomena. Even the reductionist theories which emerged later, inherited a major part of their ideas and techniques from Pasteur. Thus the development of biological science can hardly be seen as a progressive reduction of more and more biological phenomena to physical and chemical theories. Pasteur helped to convert important branches of the study of nature *from* physico-chemical *into* biological disciplines. This of course happened a hundred years ago, but it was well after the birth of modern science, and it is hard to see why the situation need be fundamentally different today. In other words there seems to be no reason to rule out the possibility of an actual *reversal* of the physical reductionist programme in biology.

I have argued that the inadequacies of Dubos's and Dagognet's accounts of Pasteur's scientific work result from a reductionist approach to *historiography* and I have tried to show how an alternative account may illuminate the role of *physical* reductionism in biology. It seems that the framework of reductionist historiography is an important barrier to a better understanding of the conflict between reductionist and antireductionist research programmes in biology.

NILS ROLL-HANSEN
Department of Philosophy
University of Oslo

REFERENCES

Bernard, C. [1878–9]: *Lecons sur les phénomènes de la vie communes aux animaux et aux végétaux.*

Berthelot, M. [1878]: 'Observation sur la note de M. Pasteur relative a la fermentation alcoolique', *Comptes Rendus de l'Académie des Sciences*, **87**, pp. 949–52. (The page reference is to the reprinting in Pasteur's *Oeuvres*, vol. **2**, pp. 586–8.)

Duclaux, E. [1896]: *Pasteur: histoire d'un esprit.* (The page reference is to the English edition: *Pasteur: the history of a mind,* 1920.)

Liebig, J. von [1871]: 'Sur la fermentation et la source de la force musculaire', *Annales de Chimie et de Physique, Quatrième série,* **23**.

Roll-Hansen, N. [1969]: 'On the Reduction of Biology to Physical Science', *Synthese,* **20**, pp. 277–89.

Schaffner, K. R. [1967]: 'Antireductionism and Molecular Biology', *Science,* **157**, pp. 644–7.

Stent, G. [1968]: 'That was the Molecular Biology that was', *Science,* **160**, pp. 390–5.

Watkins, J. W. N. [1958]: 'Confirmable and Influential Metaphysics', *Mind,* **67**, pp. 344–65.

Brit. J. Phil. Sci. **23** (1972), 363–369 *Printed in Great Britain*

Reviews

STEEN, S. W. P. [1972]: *Mathematical Logic: with Special Reference to the Natural Numbers*. Cambridge: Cambridge University Press. £15. Pp. xiii + 638.

This book is an introduction to mathematical logic and the formal systems of arithmetic. It is extremely long—almost 640 pages—but despite its length the author does not cover a great deal of territory, preferring instead to discuss a relatively small number of topics in exhaustive detail.

In *Chapter 1* the author sets up a class of basic formal systems in which to formalise mathematical arguments. Following Carnap, the author conceives of a formal system as an ordered quadruple $\mathscr{L} = \langle \mathscr{S}, \mathscr{F}, \mathscr{A}, \mathscr{P} \rangle$, where $\mathscr{S}$ is a list of *formal symbols*, $\mathscr{F}$ is a set of *formation rules* for constructing well-formed formulas from the symbols, $\mathscr{A}$ is a prescribed family of formulas called *axioms* and $\mathscr{P}$ is a description of *procedural rules* for constructing formal proofs within $\mathscr{L}$. Special sorts of formal system—e.g. the propositional and predicate calculi—constitute the primary objects of attention.

In *Chapter 2* we come to propositional calculi, in particular, the classical propositional calculus $\mathscr{P}_c$, which is set up in a Gentzen-style formulation using parenthesis-free notation for formulas. The standard metatheorems for $\mathscr{P}_c$ are proved: completeness, consistency and decidability.

Chapter 3 is devoted to predicate calculi. The usual axiom system and rules of inference are given for first-order predicate calculus $\mathscr{F}_c$, again in a Gentzen-style formulation. Herbrand's disjunction theorem is established, and then applied to yield the classical Gödel completeness theorem for $\mathscr{F}_c$—a sentence of $\mathscr{F}_c$ is provable iff it is valid in the domain of natural numbers. After this the author considers special cases of the *decision problem*, that is, to determine for a given sentence of $\mathscr{F}_c$, whether or not it is a tautology. In a later section of the chapter the author introduces semantic tableaux and proves the well-known *Craig interpolation lemma* if σ and τ are sentences of $\mathscr{F}_c$ such that $\sigma \rightarrow \tau$ is provable, then there is some sentence θ of $\mathscr{F}_c$ containing only predicate symbols common to σ and τ such that both $\sigma \rightarrow \theta$ and $\theta \rightarrow \tau$ are provable. The chapter continues with a discussion of how to formalize the von Neumann–Gödel–Bernays set theory in $\mathscr{F}_c$. One rather misleading remark occurs in *Section 3.30* in connection with the relationship between the axiom of choice and Hilbert's ϵ-symbol. In *Prop 35* of this section the author shows that the ϵ-symbol is eliminable in set theory with the axiom of choice, and asserts that this shows in effect that the axiom of choice is consistent with the other axioms of set theory. Although it is of course the case that A.C. is consistent with the other axioms (assuming that these are themselves consistent), *Proposition 35* does not prove this. What it in fact essentially shows is that if set theory with A.C. is consistent, then one can add the rule of introduction of the ϵ-symbol:

$$\frac{\exists x \phi(x)}{\phi(\epsilon x \phi(x))}$$

and allow ϵ-formulas to occur as substitution instances in the comprehension schema, without destroying that consistency.

At the end of *Chapter 3* there is a discussion of the problem of defining truth for set-theoretic statements, which leads to a brief sketch of the notion of a Boolean-valued model of set theory.

In *Chapter 4* the author constructs a weak system of free variable arithmetic A_{00}. The basic objects and operations of A_{00} consist of the natural number zero, the successor function, the operation of repeatedly applying a function, the operation of forming a function by abstraction (using λ-terms), equality and inequality between numerical expressions, and the logical connectives of disjunction and conjunction. This is a very restricted system of arithmetic in that it lacks quantifiers and so checking the truth of any of its statements does not involve a search through the infinite set of all natural numbers. And indeed, if one sets up a truth definition for A_{00} by interpreting its statements in the domain of natural numbers, it can be shown that the system is decidable with respect to this truth definition. Moreover, a system of axioms and rules of inference for A_{00} can be set up in such a way that the class of provable statements coincides with the class of true statements. Thus A_{00} is a *complete* and *decidable* arithmetical system.

Chapter 5 deals with the so-called *primitive recursive functions*, which are the number theoretic functions built up by applying substitution and simple recursion to the successor function, the constant functions, and the identity function. A simple argument shows that these functions are precisely those which can be defined in A_{00}. After a discussion of the technique of Gödel numbering, the author establishes the existence of an effectively calculable function which is not primitive recursive. In fact, let f be the function defined by setting $f(n) = 1$ if n is the Gödel number of a true sentence in A_{00}, and $f(n) = 0$ otherwise. Then f is an effectively calculable since A_{00} is decidable; a diagonal argument shows that it is not primitive recursive.

In *Chapter 6* the author adds the existential quantifier to the system A_{00}, thus obtaining a somewhat stronger system A_0. This system turns out to be adequate for representing the partial recursive functions, that is, the class of functions obtained by adding the least number operator to the defining operations for the primitive recursive functions. Again one can set up a truth definition for the sentences of A_0 by interpreting in the domain of natural numbers, and it is easy to show that the system is undecidable with respect to this truth definition. (A *false* existential statement of arithmetic might require an infinite search through the natural numbers to demonstrate its falsehood.) However, one can formulate a system of axioms and rules of inference for A_0 in such a way that the class of *provable* sentences coincides with the class of *true* sentences. Thus A_0 is a *complete* and *undecidable* arithmetical system.

Chapter 7, easily the most lucid of the book, is devoted to recursive function theory, including Turing machines, the recursion theorem, and degrees of unsolvability.

One of the results established in *Chapter 6* is the undefinability of the universal quantifier in the system A_0; in *Chapter 8* A_0 is extended to a system A which includes the universal quantifier. A truth definition may be set up for A in the same way as for A_0; A is then undecidable with respect to this truth definition. Moreover, Tarski's theorem on undefinability of truth implies that *no* recursive system of axioms and rules of inference can be complete with respect to this truth definition. Thus A is *incomplete*.

In *Chapter 9* the author discusses the class of A-definable sets of natural numbers, and introduces the arithmetical hierarchy, i.e. the natural classification of first-order definable sets of natural numbers in terms of the complexity of their defining formulas.

Chapter 10 is devoted to the properties of the system A_I which is obtained by adding to A the following induction principle: (*) infer $\forall x \phi(x)$ whenever we are given a primitive recursive function f such that, for each n, $f(n)$ is the Gödel number of a proof of $\phi(n)$. ((*) is sometimes called the constructive ω-rule.) The main result of the chapter is Gödel's second incompleteness theorem for A_I: if A_I is consistent, then the A_I-statement which asserts this consistency cannot be proved in A_I. The author gives a very detailed proof of this result, following Hilbert–Bernays' classic treatment.

Chapter 11 discusses higher order arithmetics and the analytic hierarchy, i.e. the natural classification of second-order definable sets of natural numbers in terms of the complexity of their defining formulas.

Chapter 12, the last of the book, is devoted to model theory. Ultraproducts, Henkin models, and Skolem's construction of a non-standard model of arithmetic are discussed. An unexpected bonus is a proof of Ryll-Nardzewski's theorem that the full induction schema for arithmetic is not equivalent to any finite number of its instances. The chapter closes with an informal discussion of Cohen's forcing method for producing independence proofs in set theory.

There are exercises and a brief historical note at the end of each chapter.

I confess that I found the book extremely heavy going. To begin with, the author's notation is far from lucid and so quite simple matters have a tendency to assume highly involved forms. For example, the use of parenthesis-free notation for formulas of the propositional and predicate calculi results in a veritable tangle of symbolism which I found very difficult to unravel.

Also, the description of formal systems in terms of types and syntactic categories in *Chapter 1* seems excessively involved if one is only going to use them to develop arithmetic. Certainly a beginner would find this approach almost unintelligible. Furthermore, although the book contains a painstaking treatment of the *syntactic* properties of the formal systems considered, their *semantic* features do not emerge at all clearly. The author's definition, in *Chapters 3* and *12*, of a model of a formal system as another formal system of a certain kind is hard to make sense of unless one is guided by the set-theoretic definition of a model as a relational structure. And yet this set-theoretic definition is not referred to explicitly at all.

Another stumbling-block for the beginner is the way in which the author has chosen to develop formal arithmetic. In my opinion it would have been better to discuss one of the easily grasped classical systems of formal arithmetic first, show that it is incomplete etc. As it is, the author first considers the very delicate systems A_{00} and A_0 and then strengthens them it to the complicated system A_I which has an *infinitary* rule of induction. I do not really see how the student with little previous acquaintance with mathematical logic can be expected to understand all these subtleties until he knows something about the classical systems.

I think the main reason why this book is unsatisfactory is that the author has adopted what amounts to an old-fashioned formalist approach to the subject, with its insistence on obscuring basically simple ideas under layers of precision. It might be maintained that one cannot have rigour and lucidity at the same time, and that the author could not have simplified his account of the subject without becoming sloppy. But this seems very doubtful to me: Kleene's *Introduction to Metamathematics*, published almost twenty years ago, covers much the same basic ground as the present volume in an incomparably clearer fashion without sacrificing one iota of rigour.

For all these reasons, I feel that I could not recommend this book to the student of mathematical logic.

J. L. BELL
London School of Economics
and Political Science

KÖRNER, S. [1969]: *What is Philosophy?* London: Allen Lane, The Penguin Press. £2.50. Pp. xiii+290. Also issued in paperback as KÖRNER, S. [1971]: *Fundamental Questions in Philosophy*. London: Allen Lane, The Penguin Press. £1.00. Pp. xiii+290.

'The future of philosophy is, like its past, inextricably bound up with the whole of man's intellectual life. Philosophy develops by giving birth to new disciplines or acting as socratic midwife, by internal criticism, by reacting to changes in its social and intellectual environment and by the independent exploration of new possibilities of thought.' This passage in Professor Körner's *What is Philosphy?* suggests that he advocates a committed approach to philosophy, philosophy 'inextricably bound up with the whole of man's intellectual life'.

However on other occasions Körner gives a stern warning that 'Philosophers should not, I believe, take sides in any spurious (*sic*) conflict between the humanities and the sciences. To do so is to betray their traditional function of exhibiting the interrelation between all aspects of human thought.' This seems to suggest that the function of philosophy is mere clarification and elucidation of ideas 'exhibiting the interrelation between all aspects of human thought'.

The two quotes admirably illustrate the fundamental split in Körner's book. This split represents a continual conflict between the committed philosopher and the expert technician. The philosopher attempts to grapple with problems related to the whole of man's intellectual, moral and social life. The technician constantly narrows the scope of his inquiry to the requirements of exactitude. In the overall battle the technician wins. *What is Philosophy?* is a tribute to Professor Körner's expertise as an accomplished and exceedingly skilful philosopher of the age of specialisation rather than a tribute to his moral and social commitment to the age of crisis.

It may be argued, and Körner directly and indirectly does so, that philosophy nowadays cannot be comprehended without the use of the formal techniques, of the special notations, and of the symbolic languages. Indeed. And yet, the fundamental question is whether the use of these techniques is necessary for the understanding of philosophy itself—its problems and dilemmas in the broad context of man's intellectual life, or whether the use of these techniques is only necessary for the understanding of the jargon of other philosophers—the kind of problems *they* have created often for the sake of employing their ingenious techniques alone. The issue is by no means trivial. The fact that professional philosophy is dominated by specialisation and technicalities does not mean that philosophy itself requires this specialisation and this regimentation. No doubt, *some* problems must be treated rigorously and symbolically in order to be treated at all. But 'some' does not mean all, or even most problems. As the motto for his book, Körner quotes the famous dictum of Aristotle: '. . . it is the mark of an educated mind to expect that amount of exactness in each kind which the nature of the particular subject admits.' By this definition contemporary academic philosophy is not practised by 'educated minds', for contemporary philosophy has alienated itself from the understanding and concerns not only of ordinary people but also of educated people. The main reason for this is its *narrowness*: its inability and unwillingness to undertake large, socially and humanly important problems; and its *technicality*: its preoccupation with technical means at the expense of the subject matter. It should be clearly seen that this narrowness and technicality support each other and indeed are caused by each other. One might even go further to suggest that the plight of contemporaty philosophy as a human and social enterprise is exactly the outcome of its undue specialisation.

Körner's book is a comprehensive attempt to place the whole scene of contemporary philosophy before the educated layman. That is to say, he does not question the existing state of philosophy, but rather accepts it as given and tries to render it lucid and intelligible to the 'outside world'. In this he succeeds admirably. For Körner's book, *What is Philosophy?*, is a compact compendium of problems and solutions of contemporary philosophy (to be more precise, we should say: of contemporary philosophy as it has been treated in the Anglo-Saxon world) rendered with crispness and exactitude which at once satisfies the expert and *can* be intelligible to the layman (if the latter takes sufficient pains to follow the rigorous reasoning Körner's book often requires).

Now, one would give an erroneous impression if one suggests that *What is Philosophy?* is a mere compendium. For in reconstructing various problems and rendering them lucid, Körner has engaged in original philosophical analysis of his own. His reconstruction of philosophical problems are examples of philosophical analysis at work. Perhaps the best parts of the book are those that examine the role of conceptual frameworks in theory formation, both in science and in philosophy, and those which examine the theoretical organisation of empirical discourse.

I shall here concentrate on one problem, namely the nature of harmony or disharmony between a scientific theory and the empirical evidence for it. Körner rightly argues that the logico-mathematical framework of any scientific theory imposes certain constraints and modifications on the propositions incorporated in it. However, he goes much further when he maintains that 'the transition from description to idealisation is a small price to pay for the enormous advantages thereby gained, namely conceptual simplicity and definiteness and the consequent efficiency of inference and calculation'. Here the expert technician in Körner has clearly spoken. For what is the 'enormous advantage gained?' The efficiency of inference and calculation. However, only two pages further, Körner, the philosopher, speaks again by insisting that the mistake of this attractively simple [logico-mathematical] account lies in regarding as empirical propositions, theories which do not accommodate empirical concepts at all 'but only more or less radically idealised substitutes for them'.

We have touched here upon one of the crucial points of the contemporary debate over the nature of physical theories. Körner convincingly demonstrates that because of the inexactness of some empirical predicates, many scientific theories cannot be adequately reformulated in terms of standard elementary logic. Instead of the simplified account, formalisable in standard elementary logic:

$$b_1 \wedge A. \vdash_L b_2 \tag{i}$$

he proposes:

$$[e_1 \wedge c_0 \wedge (e_1 \approx b_1) \wedge (b_1 \wedge A. \vdash_L b_2) \wedge (b_2 \approx e_2) \rightarrow e_2] \tag{ii}$$

Schema (*i*) maintains that if e_2 does not happen, then the theory T_0 has been invalidated. Schema (*ii*) on the other hand, leaves us with two possibilities: one is to abandon the theory as invalidated, and the other is to ascribe the non-occurrence of e_2 'to a failure in satisfying the conditions of identification, in particular to having neglected features of the situation described by e_1 which, in the context of employing T_0 were not in fact negligible.'

However, what it finally boils down to in Körner's language is that schema (*i*) is embedded in the logical framework of standard elementary logic, whereas schema (*ii*) is embedded in the logical framework of a three-valued logic. The operative phrase here is 'is embedded in the logical framework of'. Throughout his account and discussion of scientific theories Körner gives the impression that the empirical adequacy of scientific theories is a matter of reconstructing them in an appropriate logical framework. Körner's handling of the notion of the logical imbeddedness is sure-handed and often innovative. But it conceals one fundamental point, namely that the logical framework *cannot* account for the empirical import of scientific theories.

If one discusses the imbeddedness of scientific theories in logical frameworks, one cannot, without distorting the nature of scientific theories, neglect discussing their imbeddedness both in an ontology and in a *Weltanschauung*. If scientific theories do describe the physical world, it is because they are imbedded in a certain ontology. Furthermore, if scientific theories do give a *coherent* account of the physical world, it is because they are imbedded in a certain *Weltanschauung*. The distinction between imbeddedness in an ontology and imbeddedness in a *Weltanschauung* is subtle, but nevertheless worth maintaining. Imbeddedness in an ontology refers simply to the stuff of which the physical world is made, or of which we contend it is made. Imbeddedness in a *Weltanschauung* refers to the organisation of this ontological stuff, our comprehension of this stuff and many other subtle aspects which make us understand the nature of the physical world as we do understand it. Imbeddedness in a *Weltanschauung* thus contains metaphysical, epistemological and perhaps even ideological and neuro-biological aspects of scientific theories. J. O. Wisdom has made some useful distinctions between these two kinds of imbeddedness, though his distinctions perhaps do not coincide with mine.

Professor Körner has nothing to say about scientific theories as imbedded in ontologies or in *Weltanschauungs*. In all fairness to him I must add that he explicitly mentions (but only just mentions) that all kinds of elements—extra-scientific and extralogical—go into the formation and the development of scientific theories. However, his overall discussion of scientific theories is such that the reader, particularly one who is not thoroughly versed in contemporary philosophy (for whom the book is intended) is bound to conclude that the empirical adequacy of scientific theories is a matter of finding an adequate logical framework within which the theory can be reconstructed and imbedded.

In sum, it seems to me that Professor's Körner's book is too original to be a good ordinary textbook, particularly for beginners in philosophy. The best way of looking at *What is Philosophy?* is to view it as bridging the chasm between professional philosophy and the educated layman who wants to know what present philosophy is about, but is puzzled by its apparent incomprehension.

HENRYK SKOLIMOWSKI
University of Michigan

Brit. J. Phil. Sci. 23 (1972), 370–373 *Printed in Great Britain*

Recent Publications

BOOKS RECEIVED

Aaron, Richard I. [1972]: *Knowing and the Function of Reason.* Oxford: Oxford University Press. £3.50. Pp. x+276.

Agassi, Joseph [1972]: *Faraday as a Natural Philosopher.* Chicago and London: Chicago University Press. £5.65. Pp. xiv+359.

Alchourrón, Carlos E. and Bulygin, Eugenio [1971]: *Normative Systems* (LEP Library of Exact Philosophy, vol. 5). Vienna and New York: Springer-Verlag. DM 69. Pp. xviii+208.

Aspden, Harold [1972]: *Modern Aether Science.* Southampton: Sabberton Publications. £2.50. Pp. 165.

Austeda, Franz [1972]: *Moderne Philosophie: Probleme – Position – Profile.* Freiburg: Rombach & Co. DM 18 paper. Pp. 231.

Bar-Hillel, Yehoshua (*ed.*) [1971]: *Pragmatics of Natural Languages.* Dordrecht: D. Reidel Publishing Company. Pp. 231.

Barnes, Barry (*ed.*) [1972]: *Sociology of Science* (Penguin Modern Sociology Readings). Harmondsworth, Middlesex: Penguin Books Ltd. 55p. Pp. 397.

Bernal, J. D. [1972]: *The Extension of Man: The History of Physics before the Modern Age.* London: Weidenfeld & Nicolson £2.95. Pp. 317.

Berofsky, Bernard [1972]: *Determinism.* Princeton: Princeton University Press; London: Oxford University Press. £4.75. Pp. ix+330.

Blalock, H. M. Jr. (*ed.*) [1972]: *Causal Models in the Social Sciences.* London: Macmillan. £6.50. Pp. xi+515.

Borsari, Raffaele [1971]: *Relativity & Motion.* Bologna: Centro Superiore di Logica e Scienze Comparate. 500 lire, Pp. 35.

Buchanan, Scott [1972]: *Truth in the Sciences.* Charlottsville: University Press of Virginia. $7.50. Pp. xxiii+177.

Buck, Roger C. and Cohen, Robert S. (*eds.*) [1971]: *PSA 1970 In Memory of Rudolf Carnap* (Boston Studies in the Philosophy of Science, 8). Dordrecht: D. Reidel Publishing Company. D.fl. 120. Pp. lxvi+614.

Bunge, M. (*ed.*) [1971]: *Problems in the Foundations of Physics* (Studies in the Foundations, Methodology and Philosophy of Science, 4). Berlin, Heidelberg and New York: Springer-Verlag. DM 54. US $15.60. Pp. 162.

Cardwell, D. S. L. [1971]: *From Watt to Clausius.* London: Heinemann. £5. Pp. ix+336.

Carnap, Rudolf and Jeffrey, Richard C. [1971]: *Studies in Inductive Logic and Probability, Volume 1.* Berkeley, Los Angeles and London: University of California Press. £5.25. Pp. vi+264.

Cassirer, Ernst [1972]: *Scientific Knowledge and the Concept of Man.* Notre Dame: Indiana Press. £4.75. Pp. xi+286.

Clark, Ronald [1972]: *Sir Edward Appleton.* (With a foreword by Sir Bernard Lovell). Oxford: Pergamon Press. £4.00. $10.75. Pp. xiii+240.

Copi, Irving M. [1971]: *The Theory of Logical Types.* London: Routledge and Kegan Paul. £1.80. Pp. ix+125.

Davidson, Donald and Harman, Gilbert (*eds.*) [1972]: *Semantics of Natural Language.* Dordrecht: D. Reidel Publishing Company. Pp. x+769.

Diamond, R. J. [1971]: *The Nonmathematical Foundations of Mathematics.* California: Nambor Press. Paperback. Pp. 104.

Dutton, John M. and Starbuck, William H. (*eds.*) [1971]: *Computer Simulation of Human Behaviour.* New York, London and Toronto: John Wiley & Sons Ltd. £8.75. Pp. vi+707.

Edwards, A. W. F. [1972]: *Likelihood.* Cambridge: Cambridge University Press. £3.80. Pp. 235.

Einstein, Albert and Infeld, Leopold [1971]: *The Evolution of Physics* (Reissue of second edition, 1961). Cambridge: Cambridge University Press. £3.60. Pp. xvi+302.

Flowers, Sir Brian [1972]: *Technology and Man* (The First Leverhulme Memorial Lecture) Liverpool: Liverpool University Press. Available (free of charge) from the Registrar, University of Liverpool. Pp. 29.

Foster, Lawrence and Swanson, J. W. (*eds.*) [1972]: *Experience & Theory*. London: Duckworth. £2.75. Pp. 160.

Foster Snow, Helen [1972]: *The Chinese Communists: Sketches and Autobiographies of the Old Guard.* Westport, Connecticut: Greenwood Publishing Co. $15.00. Pp. xxi+398.

Foucault, Michel [1971]: *The Archaeology of Knowledge.* (Translated from the French by A. M. Sheridan Smith: World of Man Series, (*ed.*) R. D. Laing). London: Tavistock Publications Ltd. £3.00. Pp. 218.

Frege, Gottlob [1971]: *On the Foundations of Geometry and Formal Theories of Arithmetic* (translated and with an introduction by Eike-Henner W. Kluge). New Haven and London: Yale University Press. £4.75 cloth. Pp. xi+163.

French, Peter J. [1972]: *John Dee: The World of an Elizabethan Magus.* London: Routledge & Kegan Paul, £3.75. Pp. 243.

Georgescu-Roegen, Nicholas [1972]: *The Entropy Law and the Economic Process.* Harvard: Harvard University Press; London: Oxford University Press. £7.50. Pp. xv+457.

Ghiselin, Michael T. [1969]: *The Triumph of the Darwinian Method.* Berkeley, Los Angeles and London: University of California Press. £3.60. Pp. 287.

Giannoni, Carlo B. [1971]: *Conventionalism in Logic.* The Hague: Mouton & Company. D.fl. 32. Pp. 157.

Gil, Fernando [1971]: *La Logique du Nom.* (Essais et Philosophie, No. 5). Paris: Éditions de l'Herne. Pp. 252.

Gross, Michael and Katz, Robert [1972]: *Non-Newtonian Calculus.* Pigeon Cove, Mass.: Lee Press. $6. Pp. viii+94.

Hanson, Norwood Russell [1969]: *Perception and Discovery: An Introduction to Scientific Inquiry.* San Francisco: Freeman, Cooper and Company. $9.75. Pp. 435.

Harré, R. [1972]: *The Philosophies of Science.* Oxford: Oxford University Press. £1.30 cloth; 60p paperback. Pp. 191.

Harré, R. and Secord, P. F. [1972]: *The Explanation of Social Behaviour.* Oxford: Basil Blackwell. £3.75. Pp. 327.

Herbrand, Jacques [1971]: *Logical Writings.* ((*ed.*) Warren D. Goldfarb). Dordrecht: D. Reidel Publishing Company. D.fl. 70. Pp. 312.

Hermann, Armin [1971]: *The Genesis of Quantum Theory (1899–1913).* Cambridge; Mass. & London: MIT Press. $8.95. Pp. ix+165.

Hermann, Arne E. [1971]: *Principles of Culture and the Crisis of Modern Civilisation.* New York: Vantage Press. $4.95. Pp. 300.

Hick, John [1972]: *Biology and the Soul.* (The 25th Arthur Stanley Eddington Memorial Lecture). Cambridge: Cambridge University Press. 30p. 95 cents. U.S.A. Pp. 29.

Hindes, Robert A. (*ed.*) [1972]: *Non-Verbal Communication.* Cambridge: Cambridge University Press. £5. Pp. xiii+453.

Hindley, J. R., Bercher, B. and Seldin, J. P. [1972]: *Introduction to Combinatory Logic* (London Mathematical Society Lecture Note Series, 7). Cambridge: Cambridge University Press. £2.00. Pp. 170.

Hobson, Arthur [1971]: *Concepts in Statistical Mechanics.* New York, London and Paris: Gordon and Breach. $17.50. £7.30. Pp. ix+172.

Householder, Fred W. [1971]: *Linguistic Speculations.* Cambridge: Cambridge University Press. £6.00. Pp. xiv+352.

Howard, Nigel [1971]: *Paradoxes of Rationality: Theory of Metagames and Political Behaviour.* Cambridge, Mass. and London: MIT Press. $12.95. £6.05. Pp. xxiii+248.

Kanitscheider, Bernulf [1971]: *Geometrie und Wirklichkeit* (Erfahrung und Denken, Band 36). Berlin: Duncker & Humblot. DM 76. Pp. 389.

Kauffman, George B. (*ed.*) [1971]: *Teaching the History of Chemistry; A Symposium,* San Francisco, California, U.S.A. Budapest: Akadémiai Kiadó. £3.25. Pp. 222.

Kline, Paul [1972]: *Fact and Fantasy in Freudian Theory* (Methuen's Manuals of Modern Psychology). London: Methuen & Co. £5.00. Pp. x+405.

Klingender, Francis [1971]: *Animals in Art and Thought (to the end of the Middle Ages)* Cambridge, Mass.: The MIT Press. $25.00. Pp. xxviii+580.

Klir, George J. (*ed.*) [1972]: *Trends in General Systems Theory.* Chichester: John Wiley & Sons Ltd. £8.00. Pp 462.

Klüver, Jürgen and Wolf, Friedrich O. [undated]: *Wissenschaftskritik und sozialistische Praxis Konsequenzen aus der Studentbewegung.* Stuttgart: Froman-Holzboog. DM 32 cloth; DM 19,80 paper. Pp. 219.

König, Gert [undated]: *Was heisst Wissenschaftstheorie?* (Düsseldorfer Philosophische Bausteine III). Düsseldorf: Philosophia Verlag. DM 8.50. Pp. 59.

Kordig, Carl R. [1971]: *The Justification of Scientific Change* (Synthese Library). New York: Humanities Press. Dordrecht-Holland: D. Reidel Publishing Company. $11.00. Pp. xiv+122.

Lane, Gilles [1971]: *L'Avenir d'une Prédiction: Notes pour une Philosophie des Sciences.* Montreal: Les Presses de l'Université du Québec. Pp. ix+164.

Larson, James L. [1971]: *Reason and Experience.* Berkeley, Los Angeles and London: University of California Press. £3.35. Pp. vii+171.

Laszlo, Ervin and Wilbur, James B. (*eds.*) [1971]: *Human values and the Mind of Man* (Current Topics of Contemporary Thought, 6). New York, London and Paris: Gordon and Breach Science Publishers. £6.25. $15. Pp. 184.

Lenk, Hanz [1972]: *Erklärung: Prognose: Planung* (Skizzen zu Brennpunktproblemen der Wissenschaftstheorie). Freiburg: Verlag Rombach (Rombach Hoschschul paperback). DM 14. Pp. 120.

Lowry, Richard [1971]: *The Evolution of Psychological Theory, 1650 to the Present.* Chicago: Aldine Publishing Company. $8.95. Pp. xviii+237.

Makai, Marie [1972]: *The Dialectics of Moral Consciousness. A Contribution to the Theory of Marxist General Ethics.* Budapest: Akadémiai Kiadó. £4.50. Pp. 298.

Mathias, Peter (*ed.*) [1972]: *Science and Society 1600–1900.* Cambridge, Cambridge University Press. £2.80. $10.50. Pp. viii+166.

Meyer, Herbert W. [1971]: *A History of Electricity and Magnetism.* Cambridge, Mass. & London: MIT Press. $10.00. Pp. xvii+325.

Michalos, Alex C. [1970]: *Improve your Reasoning.* Englewood Cliffs, N.J.: Prentice-Hall, $4.95. Pp. xv+127.

Mischel, Theodore (*ed.*) (1971]: *Cognitive Development & Epistemology.* New York and London: Academic Press. $16.50. £5.70. Pp. xv+423.

Mulkay, M. J. [1972]: *The Social Process of Innovation* (Studies in Sociology Series). London and Basingstoke: Macmillan. 60p. Pp. 67.

Munitz, Milton K. (*ed.*) [1971]: *Identity and Individuation.* New York: N.Y. University Press. $9.50. Pp 261.

Needham, D. M. [1971]: *Machina Carnis.* Cambridge: Cambridge University Press. £18.00. Pp. xvi+782.

Negley, Glenn [1971]: *The Quest for Utopia.* College Park, Maryland: McGrath Publishing Company. $28.00. Pp. 599.

Nordenstam, Tore [1972]: *Empiricism and the Analytic-Synthetic Distinction.* Oslo-Hergen-Tromso: Universitetsforlaget. N.kr. 32.00. Pp. 93.

Nye, M. J. [1972]: *Molecular Reality* (History of Science Library). London: Macdonald; New York: American Elsevier Inc. D.fl. 50.00 (ca $14.75). Pp. xi+201.

Paler, Robert (*ed.*) [1971]: *The Annus Mirabilis of Sir Isaac Newton.* Cambridge, Mass. and London: MIT Press. £7.00. Pp. viii+351.

Paiget, Jean [1972]: *The Principles of Genetic Epistemology.* (Translated from the French by Wolfe Mays.) London: Routledge & Kegan Paul. £1.50. Pp. v+98.

Paiget, Jean [1972]: *Insights and Illusions of Philosophy.* (Translated from the French by Wolfe Mays.) London: Routledge & Kegan Paul. £2.50. Pp. xvii+237.

Pippard, A. B. [1972): *Reconciling Physics with Reality.* Cambridge: Cambridge University Press. 40p. £1.45. Pp. 40.

Quine, W. V. O. [1972]: *Logique Elémentaire.* Paris: Armand Colin. 11,00F. Pp. 224.

Reichenbach, Hans [1971, 1972–London]: *The Direction of Time* (edited by Maria Reichenbach). Berkeley, Los Angeles and London: University of California Press. £1.90. Pp. xi+280.

Ramsey, I. T. and Porter, Ruth (*eds.*) [1971]: *Personality and Science. An Interdisciplinary Discussion.* (A Ciba Foundation Blueprint). Edinburgh and London: Churchill Livingstone. £1.00 paperback. Pp. ix+158.

Rist, J. M. [1972]: *Epicurus: an Introduction.* Cambridge: Cambridge University Press. £3.60. Pp. xiv+185.

Rose, Alan [1971]: *Computer Logic.* London: Wiley—Interscience. £4.00. Pp.xii+180.

Ross, James F. (*ed.*) [1971]: *Inquiries into Medieval Philosophy: A Collection in Honor of Francis P. Clarke.* Westport, Conn.: Greenwood Publishing Company. $15.00. Pp. xii+329.

Rousseau, G. S. (*ed.*) [1972]: *Organic Form.* London and Boston: Routledge & Kegan Paul. £2.00. Pp. xii+108.

Russel, Robert R. [1972]: *Critical Studies in Antebellum Sectionalism: Essays in American Political and Economic History.* Westport, Connecticut: Greenwood Publishing Company. $10.00. Pp. xix+223.

Sachs, Mendel [1971]: *The Search for a Theory of Matter* (History of Science Series). New York: McGraw-Hill Book Company. $6.95. Pp. 221.

Sachs. Lothar [1972]: *Statistische Auswertungsmethoden.* Berlin, Heidelberg, New York: Springer-Verlag. DM 58.; US $17.70. Pp. xx+548.

Schaffner, Kenneth F. [1972]: *Nineteenth-Century Aether Theories.* (Selected Readings in Physics). Oxford, New York, Toronto, Sydney, Braunschweig: Pergamon Press. Pp. ix+278.

Schlegel, Richard [1972]: *Inquiry into Science.* New York: Doubleday & Company, Inc. $4.95. Pp. 128.

Shibles, Warren A. [1971]: *Metaphor: An Annotated Bibliography and History.* Wisconsin: The Language Press. $12.50. Pp. 410.

Shibles, Warren A. [1971]: *Analysis of Metaphor in the Light of W. M. Urban's Theories.* The Hague: Mouton. D.fl. 18. Pp. 171.

Simon, Michael A. [1971]: *The Matter of Life. Philosophical Problems of Biology.* New Haven and London: Yale University Press. £3.75 cloth. Pp. xi+258.

Steinberg, Danny D. and Jakobovits, Leon A. (*eds.*) [1971]: *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology.* Cambridge: Cambridge University Press. £6.80. Pp. x+603.

Stenius, Erik [1972]: *Critical Essays.* Amsterdam: North-Holland Publishing Company. D.fl. 42.

Themerson, Stefan [1972]: *Factor T.* London: Gaberbocchus Press Ltd. 80p. Pp. 63.

Thiel, Christian [1972]: *Grundlagenkrise und Grundlagenstreit.* Meisenheim: Verlag Anton Hain. DM 41,50. Pp. x+226.

Thrower, James [1971]: *A Short History of Western Atheism.* London: Pemberton Books. £1.40 cloth; 60p paper. Pp. 143.

Todd, William [1972]: *History as Applied Science: A Philosophical Study.* Detroit: Wayne State University Press. $11.95. Pp. 250.

Vorzimmer, Peter J. [1972]: *Charles Darwin: The Years of Controversy.* London: University of London Press Ltd. £4.40. Pp. xix+300.

Watson, Patty Jo; Leblanc, Steven A., Redman, Charles L. [1972]: *Explanation in Archeology: An Explicitly Scientific Approach.* New York and London: Columbia University Press. £3.30. Pp. xviii+191.

Wilson, J. R. S. [1972]: *Emotion and Object.* Cambridge: Cambridge University Press. £2.80. Pp. viii+192.

Wolstenholme, G. E. W., O'Connor, D. and Maeve, R. (*eds.*) [1971]: *Civilisation and Science: In Conflict or Collaboration?* (CIBA Foundation Publication). Amsterdam: Associated Science Publishers. £3.50. Pp. viii+227.

Woolhouse, R. S. [1971]: *Locke's Philosophy of Science and Knowledge.* Oxford: Basil Blackwell. £2.75. Pp. xi+204.

Yourgrau, Wolfgang and van der Merwe, Alwyn (*eds.*) [1971]: *Perspectives in Quantum Theory: Essays in Honor of Alfred Landé.* Cambridge, Mass. and London: MIT Press. £8.15. Pp. xxxvii+275.

Zaripov Dr. R. Kh. [1971]: *Kibernetika i Muzika.* Moscow: Akademia Nauk SSSR. 85 Kopeks. Pp. 236.

Zeman, J. (*ed.*) [1971]: *Time in Science and Philosophy: An International Study of Some Current Problems.* Amsterdam, London and New York: Elsevier Publishing Company. D.fl. 55. Pp. 305.

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

PROGRAMME OF MEETINGS 1972–73

1972

22–24 September
Annual Conference at Bristol University

9 October
Mr. D. Stove: Logical Probability and Intuition

13 November
Professor Sir Edward Bullard: Sources of the Recent Revolution in Geology

11 December
Professor R. Bernstein: The Social Sciences: Confusion and Criticism

1973

15 January
Professor Dana Scott: Does Many-Valued Logic Make Sense?

12 February
Professor A. Birnbaum: Mendelian Genetics: A Case Study in the Structure of Science

12 March
4.15 p.m.: ANNUAL GENERAL MEETING
5.15 p.m. Mr. J. Mackie: Causal Necessity

30 April
Rev. Dr. J. Russell: Chance, Necessity and Teleology

21 May
Mr. J. Worrall: Thomas Young and the 'Refutation' of Newtonian Optics

18 June
Professor I. Levi: Truth, Fallibility and the Growth of Knowledge

7–9 September
Annual Conference at Edinburgh University

Meetings are held in the Haldane Room (previously called the Joint Staff Common Room), at University College London. Meetings begin at 5.15 p.m. Tea is served at 4.45 p.m.

THE CENTER FOR THE PHILOSOPHY OF SCIENCE, AND THE DEPARTMENT OF HISTORY AND PHILOSOPHY OF SCIENCE AT THE UNIVERSITY OF PITTSBURGH

Science and Philosophical Materialism in the Soviet Union
Loren R. Graham, Department of History, Columbia University
Thursday, 28 September, 1972

Late Medieval Science and Late Medieval Philosophy: The Significance of their Effective Equivalence
John E. Murdoch, Department of the History of Science, Harvard University
Monday, 16 October, 1972

Explanatory Ideals and Practices in Current Evolutionary Theory
M. J. S. Hodge, Department of History and Philosophy of Science, University of Pittsburgh
Thursday, 2 November, 1972

The Logic and Epistemology of Statistical Inference
Ronald N. Giere, Department of History and Philosophy of Science, Indiana University, Bloomington, Ind.
Thursday, 16 November, 1972

Are There Inductive Explanations?
J. Alberto Coffa, Department of History and Philosophy of Science, Indiana University, Bloomington, Ind.
Thursday, 11 January, 1973

The Empiricist Construal of Scientific Theories; Its Rationale and Its Difficulties
Carl G. Hempel, Department of Philosophy, Princeton University
Thursday, 22 February, 1973

Scientific Method 1975: A Survey of Developments
Paul K. Feyerabend, Department of Philosophy, University of California, Berkeley
Monday, 26 March, 1973

All the meetings at 223 Langley Hall, 8.15 p.m. Inquiries can be made by writing to the Director of the Centre, Professor Adolf Grünbaum, 2509 Cathedral of Learning, University of Pittsburgh, Pittsburgh, Pa 15213.

BOSTON COLLOQUIUM FOR THE PHILOSOPHY OF SCIENCE

THIRTEENTH ANNUAL PROGRAMME, 1972–1973

Confirmation and Relevance
Wesley Salmon, Indiana University
Commentator: James Thomson, MIT
Tuesday, 10 October 1972

Toward a Rational Philosophical Anthropology
Joseph Agassi, Boston University and Tel Aviv University
Commentator: To be announced
Tuesday, 17 October 1972

Teleological and Teleonomic, A New Analysis
Ernst Mayr, Harvard University
Commentators: Ronald Munson, University of Missouri at St. Louis
Hilde Hein, College of the Holy Cross (Worcester)
Tuesday, 24 October 1972

On the Use of Projective Geometry in Physics
PAUL DIRAC
Commentator: ROGER PENROSE
Monday, 30 October 1972

Metaphysics vs. Experiment: The Dynamics of Theory Selection
STANLEY GOLDBERG, Hampshire College
Commentators: MILIC CAPEK, Boston University
MILLER BROWN, Trinity College (Hartford)
Tuesday, 14 November 1972

Domestic Causes and Purposes of Wars: A Methodological Inquiry
ARNO J. MAYER, Princeton University
Commentators: JOSEPH BOSKIN, Boston University
LOUIS O. MINK, Wesleyan University
Tuesday, 21 November 1972

The Cartesian Circle: Descartes' Defense of Reason as a Valid Response to 17th Century Scepticism
EDMOND CURLEY, Australian National University and Johns Hopkins University
Commentator: HARRY G. FRANKFURT, The Rockefeller University
Tuesday, 28 November 1972

A New Look at Phenomenology and the 'Idea' of Physics
FRANCIS J. ZUCKER, Max-Planck-Institut (Starnberg)
Commentator: PATRICK HEELAN, SUNY, Stony Brook
Tuesday, 12 December 1973

Round Table: Quantum Mechanics and Common Sense
MICHAEL R. GARDNER, Mount Holyoke College
Commentators: ARMAND SIEGEL, Boston University
MICHAEL HORNE, Stonehill College
E. P. GROSS, Brandeis University
Tuesday, 16 January 1972

Explanation and Evolution
MARJORIE GRENE, University of California, Davis and Boston University
Commentators: KENNETH SCHAFFNER, University of Pittsburgh
LYNN S. MARGULIS, Boston University
Tuesday, 23 January 1973

On the Relationship between Quantum Mechanics and Mathematics
PAUL BENIOFF, Argonne National Laboratory
Commentator: HILARY PUTNAM, Harvard University
Tuesday, 6 February 1973

Round Table: Science and Philosophical Materialism in the Soviet Union
LOREN GRAHAM, Columbia University
Commentators: THOMAS BLAKELEY, Boston College
DIRK STRUIK, MIT
MARX WARTOFSKY, Boston University
ROBERT S. COHEN, Boston University
WILLIAM MCBRIDE, Yale University
Tuesday, 13 February 1973

Fresnel Drag and the Principle of Relativity
RONALD NEWBURGH, Microwave Physics Laboratory, AFCRL—Bedford
Commentators: HUSEYIN YILMAZ, Perception Technology Corporation
ARTHUR I. MILLER, Harvard University and Lowell Technological Institute

Tuesday, 27 February 1973
Round Table: Research Strategies vs. The Relativity of Science
NORETTA KOERTGE, Indiana University
Commentators: MARX WARTOFSKY, Boston University
PHILIP QUINN, Brown University
SHELDON KRIMSKY, University of South Florida

Tuesday, 13 March 1973
Against Method
PAUL FEYERABEND, University of California, Berkeley
Commentator: IMRE LAKATOS, London School of Economics and Boston University

Tuesday, 20 March 1973
Quantum Theory of Measurement
GEN-ICHIRO NAGASAKA, Nanzan University (Nagoya) and Boston University
Commentators: JOHN STACHEL, Boston University
PAUL ROMAN, Boston University

Tuesday, 27 March 1973
Round Table: The Works of God's Hand: The Books of His Revelations, and the Machinery of Nature: 11th–17th Centuries
BENJAMIN NELSON, New School of Social Research
Commentators: JOHN MURDOCH, Harvard University
ANN BRINKLEY, Boston University
A. I. SABRA, Harvard University

Tuesday, 3 April 1973
Newton's Relational Theory of Space and Time
ARNOLD KOSLOW, CUNY, Brooklyn College
Commentator: HOWARD STEIN, Case-Western Reserve University

Tuesday, 10 April 1973
On Systematic Realism, with a Particular Application to non-Conventionalism in Physics
CLIFFORD HOOKER, University of Western Ontario
Commentators: LASZLO TISZA, MIT
ROBERT S. COHEN, Boston University

Tuesday, 17 April 1973
Space-Time Indeterminacies and Space-Time Structure
CLARK GLYMOUR, Princeton University
Commentator: JOHN GRAVES, MIT

Tuesday, 1 May 1973
Reproduction, Translation and Social Science
RICHARD RUDNER, Washington University
Commentator: THOMAS MCCARTHY, Boston University

Tuesday, 8 May 1973

Regular meetings (except Round Tables) to be held at 7.45 p.m. in Room 314 of the Sherman Union, 775 Commonwealth Avenue, at Boston University Bridge. Participation in the Round Tables is by invitation only.

For further information consult Professor Robert S. Cohen, Department of Physics, or Professor Marx Wartofsky, Department of Philosophy, Boston University.

British Journal for the Philosophy of Science

is one of 48 learned journals published by Cambridge University Press.

A descriptive catalogue of Cambridge journals with details of back volumes is available from the London and New York offices of the Press.

Cambridge journals may be ordered from a bookseller or direct from the publishers.

Cambridge University Press
Bentley House, 200 Euston Road
London, NW1 2DB
American Branch:
32 East 57th Street
New York, N.Y. 10022

Inquiry and Decision

John O'Shaughnessy details the processes involved in inquiry and the role they play in decision-making, using concepts and techniques from philosophy of science, statistics and social science.

£3.80: paperback £2.00

GEORGE ALLEN & UNWIN

RATIO

edited by Stephan Körner

VOLUME XIV	DECEMBER 1972	No. 1
RODERICK M. CHISHOLM and ROBERT G. KEIM	A System of Epistemic Logic	
K. E. TRANØY	'Ought' Implies 'Can': A Bridge from Fact to Norm	
JAN SRZEDNICKI	The Transcendental Impossibility of Solipsism	
R. S. WOOLHOUSE	From Conceivability to Possibility	
WOLFGANG CARL and ROLF P. HORSTMANN	Knowing and Claiming	
DONALD SIEVERT	Does Prichard's Moral Philosophy Rest on a Confusion	
KEITH GRAHAM	Ifs, Cans and Dispositions	
REVIEWS:		
R. M. HARE	*The Object of Morality* (G. J. Warnock)	
W. H. WALSH	*Early German Philosophy—Kant and his Predecessors* (Lewis White Beck)	
EDGAR MORSCHER	*Rechtslogik. Versuch einer Anwendung moderner Logik auf das Juristische Denken* (Ota Weinberger)	

Price £1.75 net **Annual Subscription: £3.00 net**

BASIL BLACKWELL · OXFORD

Studies in History and Philosophy of Science

Vol. 3 No. 3 November 1972

The three essays in this issue develop certain methodological themes which arise from a study of the writings of Johannes Kepler (1571–1630). The papers were presented at an international symposium held at Weil der Stadt, Kepler's birthplace, during August 1971.

The contents will include:

JURGEN MITTELSTRASS *Methodological Elements of Keplerian Astronomy*

ROBERT S. WESTMAN *Kepler's Theory of Hypothesis and the 'Realist Dilemma,*

GERD BUCHDAHL *Methodological Aspects of Kepler's Theory of Refraction*

The journal is published quarterly by Macmillan Journals Ltd.

Annual subscription to all countries £5.00, except USA and Canada $15.00 (by airfreight). Subscriptions may be ordered by sending the appropriate remittance to: Macmillan Journals Ltd., Subscription Department, Brunel Road, Basingstoke, Hampshire, England.

FOUNDATIONS OF LANGUAGE

International Journal of Language and Philosophy

Board of Editors: Morris Halle (M.I.T.), Peter Hartman (Konstanz), K. Kunjunni Raja (Madras), Benson Mates (Univ. of California), J. F. Staal (Univ. of California), Pieter A. Verburg (Groningen), John W. M. Verhaar (Djakarta).

Editorial Assistants: W. G. Klooster and H. J. Verkuyl.

CONTENTS OF VOLUME 8 NO. 3

Asa Kasher / Sentences and Utterances Reconsidered—Saul Levin / Non-Paradigmatic Forms: Suppletion or Preemption?—Ivan Lowe / On the Relation of Formal to Sememic Matrices with Illustrations from Nambiquara—Robert Litteral / Rhetorical Predicates and Time Topology in Anggor.

Discussion: The Limits of Intuition (Thomas G. Bever).

Reviews: Marcia L. Colish / *The Mirror of Language: A Study in the Medieval Theory of Knowledge* (W. Keith Percival)—Edward H. Bendix / *Componential Analysis of General Vocabulary: the Semantic Structure of a Set of Verbs in English, Hindi, and Japanese* (Robert I. Binnick)—J. Dubois, F. Edeline, J. M. Klinkenberg, P. Minguet, F. Pire, H. Trinon (eds.) / *Rhétorique générale, by le group μ* (Seymour Chatman)—Geoffrey Wagner / *On the Wisdom of Words* (W. Bronzwaer)—Karl Heinz Wagner / *Generative Grammatical Studies in the Old English Language* (Vol. XI, *Wissenschaftliche Bibliothek*) (Herbert E. Brekle).

Subscription price per volume of four issues Dfl. 115 (US $36.80), including postage. About six or seven issues are published yearly. Personal subscription price on request.

D. REIDEL PUBLISHING COMPANY / P.O. Box 17 - DORDRECHT - HOLLAND

Noûs

Nihil philosophici a nobis alienum putamus

A Quarterly Journal of Philosophy
edited by Hector-Neri Castañeda
Philosophy Department, Indiana University
Bloomington, Indiana 47401

NOÛS is committed to the publication of first-rate essays and short discussions on philosophical problems regardless of the author's philosophical school or point of view. It includes critical studies of recently published books rather than reviews and critical summary surveys of studies as well as symposia dealing with specific philosophical problems.

A quarterly, NOÛS is published in March, May, September, and November each year. For *individuals* the subscription rates are: one year, $8.00; two years, $15.00; three years, $22.00. For *bona fide students*: $5.00 per year. For *institutions*: $12.00 per year. Add 45 cents per year for foreign subscriptions.

Volume VI, Number 3, September 1972

CONTENTS

RICHARD T. GARNER—On Saying What is True
LEONARD LINSKY—Two Concepts of Quantification
RUTH BARCAN MARCUS—Quantification and Ontology
J. WILLIAM FORGIE—Frege's Objection to the Ontological Argument
SHELDON KRIMSKY—The Multiple-World Thought Experiment and Absolute Space
JAAKKO HINTIKKA—Transcendental Arguments: Genuine and Spurious
C. G. PRADO—Internationality and Causal Analysis

Order from
The Secretary, NOÛS
Department of Philosophy
Indiana University
Bloomington, Indiana 47401

Revue Internationale de Philosophie

Vingt-sixième année No. 99/100 1972, Fasc. 1–2

Americana

SOMMAIRE

Edouard Morot-Sir, *L'Amérique et le besoin philosophique*
Gérard Deledalle, *Présence du Pragmatisme*
Manley Thompson, *Quine and the Inscrutability of Reference*
N. Mouloud, *Quine et la critique de la proposition*
Morris Weitz, *Open Concepts*
Conrad J. Koehler, *Quine on the Ontological Controversies*
Leonard Linsky, *Referential Opacity*
Andrew J. Reck, *Contemporary American Speculative Philosophy*
John E. Smith, *History of Science and the Ideal of Scientific Objectivity*

VARIÉTÉ ANALYSES ET COMPTES RENDUS OUVRAGES REÇUS
REVUE DES REVUES VIE PHILOSOPHIQUE

Direction Scientifique:
Professeurs Ph. Devaux, H. Gouhier
et Ch. Perelman.

Abonnements: Belgique: ...... 450 FB.
Pays Européens: .. 475 FB.
Autres Pays: .... 500 FB.

Administration:
Mme Simone Servais
avenue de l'Université 99
1050 Bruxelles—Belgique
Tel.: 02/ 48 03 70.

Derniers Numéros Parus:
1971: N°95/96. Hypothèse et Confirmation. N°97. L'Egalité.
N°98. Logique et Méthodologie des Sciences en U.R.S.S.

Numéros Prévus:
Feuerbach. Russell. Meinong.
Etudes philosophiques et Informatique. Lukàcs.

SYNTHESE

An International Journal for Epistemology, Methodology and Philosophy of Science

Editor-in-Chief: JAAKKO HINTIKKA
Academy of Finland and Stanford University

Review Editor: Dagfinn Føllesdal, *University of Oslo*
Associate Editor: Risto Hilpinen, *University of Turku*
Editorial Assistant: Juha Manninen, *Helsinki*

Contents of Volume 23, No. 4

Stefan Nowak / Inductive Inconsistencies and Conditional Laws of Science—Richard T. Hull / Feyerabend's Attack on Observation Sentences—Robert Howell / Seeing As—Marshall Swain / An Alternative Analysis of Knowing—Richard Rorty / Indeterminacy of Translation and of Truth—Gunnar Niemi / On the Existence of a Model Antinomy.

Discussions: Can Language Be Explained Functionally? (Geoffrey Sampson)—Reply (Stephen Toulmin).

Reviews: Y. Bar-Hillel (ed.), Mathematical Logic and Foundations of Set Theory (F. R. Drake)—Wolfgang Müller-Lauter, Nietzsche, Seine Philosophie der Gegensätze und die Gegensätze seiner Philosophie (Trond Berg Eriksen).

Communications of the International Union of History and Philosophy of Science Division of Logic, Methodology, and Philosophy of Science—Boston Colloquium for the Philosophy of Science (1971–1972)—Books Received—Index of Names.

Subscription price per volume of four issues Dfl. 90—[US $28.80] including postage. Two volumes will be published yearly. Personal subscription price on request.

D. REIDEL PUBLISHING COMPANY
P.O. Box 17 / DORDRECHT-HOLLAND

THE

REVIEW OF METAPHYSICS

A PHILOSOPHICAL QUARTERLY

Edited by JUDE P. DOUGHERTY

Volume XXV Supplement in Honour of Paul Weiss Spring 1972
on his 70th Birthday
and the 25th Anniversary of the Founding of the *Review*
Edited by Richard J. Bernstein

Andrew J. Reck	*Paul Weiss's Concept of Being*
Marc F. Griesbach	*The Epistemology of Paul Weiss*
Robert Neville	*Knowledge and Being: Comments on Griesbach and Reck*
Paul G. Kuntz	*The Achievement of Paul Weiss*
Kenneth L. Schmitz	*Art and Existence: Reflections on Paul Weiss's Modal Philosophy of Art*
Nathan Rotenstreich	*The Ontological and Epistemological Dimensions of History*
Charles Hartshorne	*Paul Weiss's* The God We Seek
Irwin C. Lieb	*Individuals and the Past*
Ellen S. Haring	*Actualities as Private and Public*
Paul Weiss	*A Response*

Bibliography of Paul Weiss's Publications

Individual Subscription $6.00 Student Subscription $4.00
Institutional Subscription $12.00

THE CATHOLIC UNIVERSITY OF AMERICA, WASHINGTON, D.C. 20017
Twenty-Year Cumulative Index, 1947-1967 is available for $3.00 a copy.

INQUIRY An Interdisciplinary Journal of Philosophy and the Social Sciences
Edited by Arne Naess
and Alastair Hannay

Number 3 *Autumn 1972* *Volume 15*

Marx, Popper, and 'Historicism'	W. A. Suchting
Wittgenstein's Blue and Brown Books (Part One)	Paul Wienpahl
Discussions:	
Of Words, Tools, and Actions	Samuel Pagee
Pre-Emptive Violence: a Reply to Gauthier	F. S. McNeilly
Kantian Intuitions	Jaakko Hintikka
Review Discussions:	
Ontology without Taste (R. Wollheim, *Art and its Objects*)	Price Charlson
Notes on the Analysis of Structure and Structuralist Ideologies (Jeanne Parain-Vial, *Le Sens du présent, Essai sur la rupture de l'unité originelle*, etc.)	André Mercier

Published quarterly by Universitetsforlaget, P.O. Box 307
Blindern, Oslo 3, Norway

Subscription: N. kr. 60.00, £3·60 $10.00, postage included

MIND

A Quarterly Review of Psychology and Philosophy
edited by Professor **DAVID HAMLYN**

OCTOBER 1972

I	Logical Possibility	G. SEDDON
II	Definitions and 'Clusters'	D. E. COOPER
III	How to Tell Your Friends from Machines	G. ROBINSON
IV	Religious Belief and Philosophical Analysis	D. A. MCDOUGALL
V	Psychology and the Philosophy of Science	G. WESTLAND
VI	On Thinking	J. HARTNACK
VII	The Intelligibility of Wants	A. J. WATTS
VIII	Discussions:	
	The Connection Argument	G. ISEMINGER
	Whether the Theory of Family Resemblances Solves the Problem of Universals	J. W. THORP
	Urmson on Evaluation from a Point of View	J. W. NICKEL
	Vagueness and Colour Predicates	C. BOYNE
	Austinian Ifs	R. J. FOGELIN
	Reply to Ruth Anna Putnam	W. T. SCOTT
	A Doubt about the Normative Theory of Belief	D. PUGMIRE
	A Solution to an Ethical Paradox	F. J. LEAVITT
	Description and Evaluation	D. CONWAY
	Hedman on Explanation	J. E. WHITE
	On a New Escape from Logical Determinism	T. CHAPMAN
	A Note on Hampshire's Analogy	J. BELFORD
IX	Critical Notes:	
	The Bounds of Sense: an Essay on Kant's Critique of Pure Reason. by P. F. Strawson	W. CERF
X	New Books	
XI	Notes	

Annual Subscription £1.25, post free; single copy 40p, plus postage
Published for MIND ASSOCIATION by
BASIL BLACKWELL 49 BROAD STREET OXFORD

NOTRE DAME JOURNAL OF FORMAL LOGIC

BOLESŁAW SOBOCIŃSKI, *Editor*

Managing Editor:
JOHN THOMAS CANTY

Business Manager:
ANN W. LAUER

Published quarterly by the University of Notre Dame in January, April, July and October of each year, the Journal is devoted primarily to logic and the foundations of mathematics (i.e. algebraic logic, classical logical systems, constructive mathematics, methodology of deductive systems, model theory, non-classical formal systems, proof theory, recursion theory, and set theory).

Articles for publication and all editorial correspondence should be sent to the *Editor, P.O. Box 28, Notre Dame, Indiana* 46556, *U.S.A.*

Subscriptions are $16.00 per year for institutions and $10.00 for individuals. Back volumes of the Journal are available. Orders for subscriptions and back volumes (accompanied by payment), and changes of address should be sent to the *Business Manager, P.O. Box 5, Notre Dame, Indiana* 46556, *U.S.A.*

Philosophy of Science

Official Journal of the Philosophy of Science Association

Editorial Board

Richard S. Rudner (Editor-in-Chief), Gustav Bergmann, May Brodbeck, Arthur Burks, C. West Churchman, Theodosius Dobzhansky, Adolf Grünbaum, Edward Madden, Henry Margenau, Gerald J. Massey, Wesley C. Salmon, Melford E. Spiro, Peter D. Asquith (Managing Editor). Books Reviews, James Van Evra.

Vol. 38 December, 1971 No. 4

CONTENTS

CARL R. KORDIG **The Comparability of Scientific Theories**

ROLF A. EBERLE **Replacing one Theory by Another under Preservation of a Given Feature**

J. N. HATTIANGADI **Alternatives and Incommensurables: The Case of Darwin and Kelvin**

MICHAEL R. GARDNER **Is Quantum Logic Really Logic?**

WILLIAM TODD **Intentions and Programs**

THOMAS NICKLES **Covering Law Explanation**

JACK C. CARLOYE **An Interpretation of Scientific Models Involving Analogies**

JOSEPH M. FIRESTONE **Remarks on Concept Formation: Theory Building and Theory Testing**

HERMAN ERLICHSON **Discussion: The Lorentz-Fitzgerald Contraction Hypothesis and the Combined Rod Contraction-Clock Retardation Hypothesis**

PHILOSOPHY OF SCIENCE is published quarterly: in March, June, September, and December. Membership dues in the Association, which cover subscriptions to the Journal, are $10.00 *per annum*. Applications for membership should be sent to PROFESSOR PETER D. ASQUITH, Philosophy Department, Michigan State University, East Lansing, Michigan 48823. The price of subscription to non-members is $15.00. Students may subscribe at the special subscription rate.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by

NICHOLAS RESCHER

VOLUME 9/NUMBER 4 OCTOBER 1972

CONTENTS

I K. W. RANKIN: *The Non-Causal Self-Fulfilment of Intention*
II BARUCH BRODY: *Locke on the Identity of Persons*
III JOHN KEKES: *Fallibilism and Rationality*
IV JOHN LACHS: *Fichte's Idealism*
V RAYMOND MARTIN: *On Weighting Causes*
VI P. L. MCKEE: *Non-Conscious Seeing*
VII ROBERT F. LADENSON: *Legitimate Authority*
VIII EDWARD QUEST: *Why Is It Difficult to Deserve Praise?*

Annual subscription: $18.00 (Institutions), $8.00 (Individuals).
Single Issues: $5.00 (Institutions), $3.00 (Individuals).

BASIL BLACKWELL
Oxford, England

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY

Edited by Marvin Farber

SEPTEMBER, 1972, VOLUME XXXIII, No. 1

CONTENTS

I ARTICLES

- Gaston Bachelard's Philosophy of Imagination: An Introduction, by EDWARD K. KAPLAN
- Conceptual Structure of Classical Logic, by JOHN CORCORAN
- Eidetic Analysis, Informal Rigor, and a Phenomenological Critique of Carnap's Notion of Explication, by ROBERT TRAGESSER
- Scientific Theory: Empirical Content, Embedded Ontology, and Weltanschauung, by J. O. WISDOM
- Reasons *vs.* Causes in Explanation of Action, by RUTH MACKLIN
- Justice, Reasonableness, and the Similar Handling of Similar Cases, by PHILLIP MONTAGUE

II DISCUSSIONS, REVIEWS

III NOTES AND NEWS, RECENT PUBLICATIONS

Subscription prices $7.50 per year for individuals.

$9.00 per year for institutions.

Separate issues, $2.00 and $2.50 respectively.

Postage and handling for back issues only.

Papers submitted for publication will not be returned unless accompanied by a self-addressed stamped envelope, or equivalent postage.

A circular listing the main contents of the Journal since 1940 will be sent upon request.

Address all communications to
Philosophy and
Phenomenological Research
State University of New York at Buffalo
Buffalo, New York 14226